

FACULTAD DE ESTUDIOS ESTADÍSTICOS

MÁSTER EN MINERÍA DE DATOS E INTELIGENCIA DE NEGOCIOS

Curso 2017/2018

Trabajo de Fin de Máster

**¿Cómo es el mercado laboral para los
egresados en el curso académico
2009/2010 en España según su trayectoria
académica?**

Alumno: Teresa Gutiérrez Quintana

***Tutores: Inés María Cáceres García
Rosario Cintas del Río***

Noviembre de 2018



UNIVERSIDAD COMPLUTENSE
MADRID

ÍNDICE

1.	Introducción	1
1.1.	Contextualización.....	1
2.	Fuente de datos.....	3
3.	Objetivos	3
4.	Metodología	3
4.1.	Recursos informáticos utilizados	3
4.2.	Técnicas descriptivas	4
4.3.	Análisis Clúster	4
4.3.1.	Selección de variables	4
4.3.2.	Análisis Clúster Jerárquico para determinar el número de clústeres	5
4.3.3.	Análisis Clúster no Jerárquico: k-modas.....	7
4.3.4.	Caracterización de los clústeres	8
4.4.	Análisis de Correspondencias Simples.....	9
4.5.	Árboles de clasificación.....	11
5.	Perfil general de la muestra y de la subpoblación de egresados.....	13
6.	Clasificación de los grupos de perfiles académicos	20
6.1.	Selección de variables.....	21
6.2.	Análisis clúster jerárquico	22
6.3.	Análisis clúster no jerárquico: método de k-modas	24
6.4.	Caracterización de los clústeres del perfil académico.....	25
6.5.	Resumen de la caracterización y descripción de los perfiles académicos	29
7.	Clasificación de los grupos de perfiles laborales.....	31
7.1.	Población que se encuentra trabajando.....	32
7.1.1.	Selección de variables	32
7.1.2.	Análisis clúster jerárquico.....	34
7.1.3.	Análisis clúster no jerárquico: método de k-modas.....	35
7.2.	Población desempleada que sí ha trabajado alguna vez.....	35
7.2.1.	Selección de variables	36
7.2.2.	Análisis clúster jerárquico.....	36
7.2.3.	Análisis clúster no jerárquico: método de <i>k-modas</i>	37
7.3.	Obtención de los clústeres totales del ámbito laboral.....	38

7.4.	Caracterización de los clústeres conformados del ámbito laboral.....	38
7.4.1.	Caracterización de la población que está trabajando	38
7.4.2.	Caracterización de la población desempleada que sí ha trabajado alguna vez	45
7.4.3.	Caracterización de los individuos desempleados que no ha trabajado nunca	50
7.4.4.	Caracterización de los individuos inactivos.....	51
7.5.	Resumen de la caracterización y descripción de los perfiles laborales.....	51
8.	Relaciones entre los perfiles académicos y laborales.....	52
8.1.	Estudio de las proporciones sobre el total de la población de cada modalidad.	53
8.2.	Contraste de independencia de clústeres y aportaciones al Chi-cuadrado	53
8.3.	Análisis con las observaciones y modalidades elegidas	55
8.4.	Representación gráfica y principales conclusiones	60
9.	Detección de factores influyentes que determinan la situación laboral	62
9.1.	Factores influyentes que determinan la presencia o ausencia de empleo	63
9.2.	Factores influyentes que determinan el tipo de contrato laboral	65
10.	Conclusiones.....	68
11.	Bibliografía.....	69
A.	Anexo.....	70
A.1.	Tablas, gráficos y material complementario.....	70
A.2.	Cuestionario EILU.....	106
A.3.	Código de SAS Base.....	159
A.4.	Código de R.....	172
A.5.	Diagrama de SAS Enterprise Miner.....	194

ÍNDICE DE TABLAS

Tabla 1. Distribución de la rama de estudio para cada situación laboral	15
Tabla 2. Resultados del contraste de independencia entre variables y situación laboral	18
Tabla 3. Afijación proporcional de la muestra estratificada para los perfiles académicos	21
Tabla 4. Resultado de las pruebas de selección de variables para perfiles académicos	22
Tabla 5. Variables seleccionadas para obtener los perfiles académicos	22
Tabla 6. Resultados de la elección del método de cálculo de distancias entre clústeres de los perfiles académicos.....	23
Tabla 7. Estadísticos para la selección del número de clústeres de los perfiles académicos	24
Tabla 8. Variables utilizadas para la formación del mapa de calor de los perfiles académicos	26
Tabla 9. Resumen de las principales aportaciones a la inercia entre la rama de estudio y los clústeres del perfil académico	28
Tabla 10. Resumen de la caracterización del perfil académico	29
Tabla 11. Principales titulaciones de los clústeres de los perfiles académicos.....	30
Tabla 12. Reagrupación de la variable "año en el que empezó a trabajar en su actual empleo"	32
Tabla 13. Reagrupación de la variable "país en el que está trabajando" el individuo...	33
Tabla 14. Resultado de las pruebas de selección de variables para perfiles laborales de individuos trabajando.....	33
Tabla 15. Variables seleccionadas para obtener los perfiles laborales de individuos trabajando	33
Tabla 16. Resultados de la elección del método de cálculo de distancias entre clústeres de los perfiles laborales de individuos trabajando.....	34
Tabla 17. Variables disponibles para los individuos en desempleo	36
Tabla 18. Resultado de las pruebas de selección de variables para perfiles laborales de individuos desempleados	36
Tabla 19. Variables seleccionadas para obtener los perfiles laborales de individuos desempleados.....	36
Tabla 20. Resultados de la elección del método de cálculo de distancias entre clústeres de los perfiles laborales de individuos desempleados.....	37
Tabla 21. Agregación de los clústeres obtenidos de todos los perfiles laborales.....	38
Tabla 22. Conjunto de variables con más de dos categorías de los perfiles laborales de individuos trabajando.....	39
Tabla 23. Calidad, masa e inercia de los clústeres del perfil laboral de individuos trabajando del ACS con la base de cotización a la Seguridad Social en 2014.....	40
Tabla 24. Resumen de las principales aportaciones a la inercia de las dimensiones de la base de cotización a la Seguridad Social en 2014 y los clústeres de los perfiles laborales de individuos trabajando.....	41

Tabla 25. Calidad, masa e inercia de los clústeres del perfil laboral de individuos trabajando del ACS con la actividad principal del trabajo actual.....	42
Tabla 26. Resumen de las principales relaciones entre la actividad principal del trabajo y los clústeres del perfil laboral de individuos trabajando	42
Tabla 27. Calidades de la representación sobre 2 y 3 ejes de las modalidades de los perfiles laborales de individuos trabajando	43
Tabla 28. Resumen de las principales relaciones entre la situación profesional y los clústeres de los perfiles laborales de individuos trabajando	44
Tabla 29. Variables utilizadas para la formación del mapa de calor de los perfiles laborales de individuos desempleados	45
Tabla 30. Variables con más de dos categorías utilizadas para la caracterización de desempleados.....	46
Tabla 31. Calidades de representación sobre los ejes de los clústeres al retener dos dimensiones.....	47
Tabla 32. Reagrupación de la variable "número de empleadores"	47
Tabla 33. Resumen de las principales contribuciones a la inercia del número de empleadores y los clústeres de los perfiles laborales de desempleados.....	48
Tabla 34. Resumen de las principales contribuciones a la inercia del tiempo que ha trabajado antes y los clústeres del perfil laboral de individuos desempleados	49
Tabla 35. Conjunto de variables utilizadas para caracterizar a individuos desempleados que no han trabajado nunca	50
Tabla 36. Conjunto de variables utilizadas para caracterizar a individuos desempleados inactivos.....	51
Tabla 37. Resumen de la caracterización del conjunto de perfiles laborales	52
Tabla 38. Modalidades con una presencia inferior al 5% sobre el total de la población y su calidad de la representación sobre el eje	53
Tabla 39. Valores observados menos esperados de las modalidades de los clústeres de los perfiles académicos y laborales	55
Tabla 40. Calidades de la representación sobre los ejes de los clústeres del perfil laboral	56
Tabla 41. Calidades de la representación sobre los ejes de los clústeres del perfil académico.....	56
Tabla 42. Cosenos al cuadrado de los clústeres del perfil laboral	57
Tabla 43. Cosenos al cuadrado de los clústeres del perfil académico	58
Tabla 44. Contribuciones parciales a la inercia de los clústeres del perfil laboral.....	58
Tabla 45. Contribuciones parciales a la inercia de los clústeres del perfil académico ..	59
Tabla 46. Resumen de los cosenos al cuadrado y las contribuciones parciales a la inercia de los clústeres del perfil académico	59
Tabla 47. Resumen de los cosenos al cuadrado y las contribuciones parciales a la inercia del perfil laboral.....	60
Tabla 48. Variables utilizadas en los árboles de clasificación	63
Tabla 49. Importancia de variables del árbol de clasificación de la situación laboral ...	64
Tabla 50. Importancia de variables del árbol de clasificación de la situación profesional	66

ÍNDICE DE GRÁFICOS

Gráfico 1. Evolución de la tasa de paro desde 2002(T1) hasta 2018(T1). Fuente: INE....	1
Gráfico 2. Evolución de la tasa de paro de personas entre 16 y 29 años desde 2002T1 hasta 2018T2. Fuente: INE.....	2
Gráfico 3. Evolución de la tasa de paro por nivel de estudios alcanzados desde 2002T1 hasta 2018T2. Fuente: INE.....	2
Gráfico 4. Distribución de la edad de los egresados	14
Gráfico 5. Distribución de la situación laboral de los egresados.....	14
Gráfico 6. Distribución del tipo de universidad en la que estudiaron los egresados.....	14
Gráfico 7. Distribución de la rama de estudio de los egresados	15
Gráfico 8. Distribución del sexo por situación laboral	16
Gráfico 9. Distribución del tipo de universidad por situación laboral.....	16
Gráfico 10. Distribución de la rama de estudio por tipo de universidad de egresados que se encuentran trabajando	17
Gráfico 11. Distribución de la rama de estudio por tipo de universidad de egresados que se encuentran desempleados.....	17
Gráfico 12. Distribución del idioma acreditado oficialmente por situación laboral	18
Gráfico 13. Distribución del tipo de contrato por sexo	19
Gráfico 14. Distribución por sexo de la base de cotización a la Seguridad Social en 2014	20
Gráfico 15. Pseudo F y Pseudo T para la selección del número de clústeres de los perfiles académicos	24
Gráfico 16. Mapa de calor para la caracterización de los perfiles académicos.	26
Gráfico 17. Descomposición de la inercia del ACS de la variable rama de estudio de los perfiles académicos	27
Gráfico 18. Representación de las modalidades sobre los ejes 1 y 2 del ACS de la rama de estudio y el perfil académico.....	28
Gráfico 19. Pseudo F y Pseudo T para la selección del número de clústeres de los perfiles laborales de individuos trabajando	35
Gráfico 20. Pseudo F y Pseudo T para la selección del número de clústeres de los perfiles laborales de individuos desempleados	37
Gráfico 21. Mapa de calor para la caracterización de los perfiles laborales de individuos trabajando	39
Gráfico 22. Descomposición de la inercia del ACS de la base de cotización a la Seguridad Social de los perfiles laborales de individuos trabajando	40
Gráfico 23. Descomposición de la inercia del ACS de la actividad principal del empleo de los perfiles laborales de individuos trabajando.....	41
Gráfico 24. Descomposición de la inercia del ACS de la variable situación profesional de los perfiles laborales de individuos trabajando.....	43
Gráfico 25. Representación de las modalidades sobre los ejes 1 y 2 del ACS de la situación profesional y los perfiles laborales de individuos trabajando.....	44
Gráfico 26. Mapa de calor para la caracterización de los perfiles laborales de individuos desempleados.....	46

Gráfico 27. Descomposición de la inercia del ACS de la variable número de empleadores de los perfiles laborales de desempleados	46
Gráfico 28. Representación de las modalidades sobre los ejes 1 y 2 del ACS del número de empleadores y los clústeres de perfiles laborales de individuos desempleados que sí han trabajado antes.....	48
Gráfico 29. Descomposición de la inercia del ACS de la variable "tiempo que ha trabajado antes" de los perfiles laborales de individuos desempleados	49
Gráfico 30. Representación de las modalidades sobre los ejes 1 y 2 del ACS del tiempo que ha trabajado antes y el perfil laboral de individuos desempleados	50
Gráfico 31. Descomposición de la inercia del ACS de clústeres de perfiles académicos y laborales	54
Gráfico 32. Mapa de calor de las aportaciones al Chi-Cuadrado de cada modalidad ...	54
Gráfico 33. Representación sobre los ejes retenidos de los clústeres del perfil académico y laboral	60
Gráfico 34. Diagrama de cajas de la tasa de clasificación errónea del árbol de clasificación de la situación laboral.....	63
Gráfico 35. Diagrama del árbol de clasificación ganador de la situación laboral	64
Gráfico 36. Diagrama de cajas de la tasa de clasificación errónea del árbol de clasificación de la situación profesional.....	65
Gráfico 37. Diagrama del árbol de clasificación ganador de la situación profesional ...	67

1. Introducción

El presente trabajo tiene como objetivo principal buscar la relación entre los aspectos académicos y laborales de los recién egresados en el curso académico 2009/2010 de universidades españolas, tanto públicas como privadas.

La investigación de este tema se realizó por el interés de conocer si los aspectos académicos influyen sobre la situación laboral, y cuáles de ellos resultan relevantes a la hora de encontrar empleo. Con esto conseguimos brindar una información útil a los futuros estudiantes.

Los datos con los que se desarrolla el trabajo son de carácter oficial y su origen es Encuesta de Inserción Laboral de Titulados Universitarios (EILU) procedente del Instituto Nacional Estadística (INE). En la EILU se consideró el conjunto de universidades del territorio español y fue realizada entre septiembre de 2014 y febrero de 2015, ya que se considera que se necesitan al menos tres años para la estabilización en el mercado laboral.

Para conseguir el objetivo propuesto, el trabajo se estructura de la siguiente manera: en primer lugar, se realizará un análisis de la evolución de la situación laboral en España con especial énfasis en la crisis de 2008 y el efecto de esta entre los jóvenes; a continuación, se explicará la metodología seguida a lo largo del trabajo indicando las fases en las que serán aplicadas; le sigue un análisis descriptivo de los datos disponibles con el fin de familiarizarnos con la población con la que se trabaja; posteriormente, se desarrollan los objetivos con las técnicas explicadas en el apartado de metodología, y por último, se indican las principales conclusiones alcanzadas.

1.1. Contextualización

A raíz de la crisis económica y financiera de 2008 en España, el impacto laboral no fue homogéneo en el conjunto de la población [1]. Estas diferencias están asociadas a variables como el sexo, la edad, nivel de formación y nacionalidad.

En concreto, la caída del empleo ha afectado más a los hombres que a las mujeres. Esto se debe a que los sectores que donde hubo más destrucción del empleo, estaban compuestos en su mayoría por hombres. A pesar de eso, el desempleo femenino durante la crisis, y actualmente, es mayor. El Gráfico 1 muestra la evolución de la tasa de paro distinguiendo por género, desde 2002 hasta el primer trimestre (T1) de 2018.

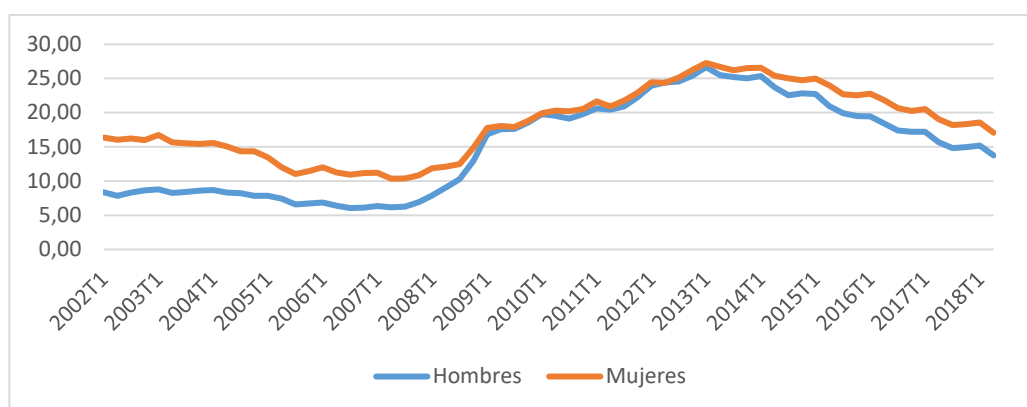


Gráfico 1. Evolución de la tasa de paro desde 2002(T1) hasta 2018(T1). Fuente: INE

Igualmente, el impacto de la crisis en el mercado laboral fue mayor en el grupo de personas jóvenes con edad de 16 a 29 años [1]. Esta vulnerabilidad del empleo juvenil, recurrente en etapas de crisis, se puede explicar en base a ciertos factores. El más destacable es la elevada precariedad laboral que existe entre los jóvenes debido a la alta temporalidad del empleo. En situaciones de crisis son estos los empleos que más se ven afectados [2].

A continuación, mostramos en el Gráfico 2 la evolución del paro de personas de 16 a 29 años, cuya fuente es el INE.

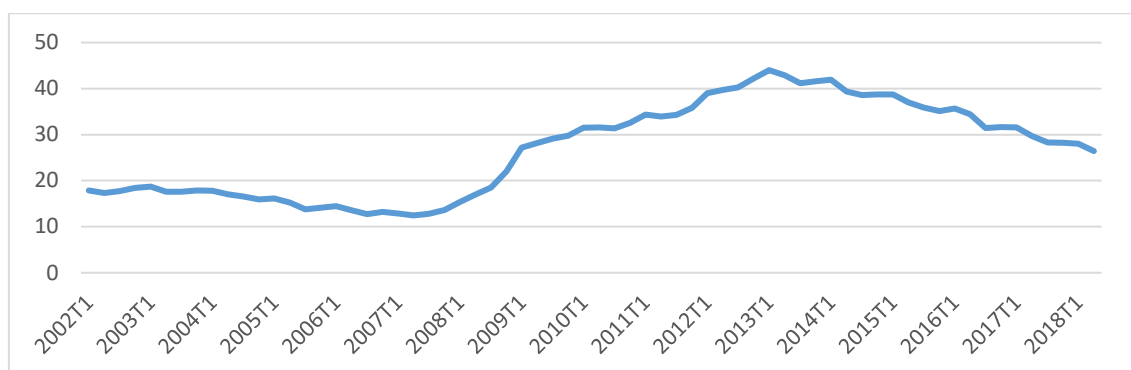


Gráfico 2. Evolución de la tasa de paro de personas entre 16 y 29 años desde 2002T1 hasta 2018T2. Fuente: INE

El nivel de formación es otra de las variables asociadas a un impacto no homogéneo de la crisis. Esta ha afectado con mayor intensidad a las personas con niveles más bajos de formación. Se debe a que estas personas ocupan puestos de trabajo con baja productividad, inestables y vinculados a las variaciones de la demanda. Por ejemplo, el sector inmobiliario da empleo a un gran número de trabajadores no cualificados, y este es uno de los sectores más afectados por la crisis.

A continuación, mostramos el Gráfico 3 con la evolución de la tasa de paro por nivel de estudios alcanzados desde 2002 hasta 2018. En ella vemos que las personas con educación superior es la que menos retroceso de trabajo ha sufrido.

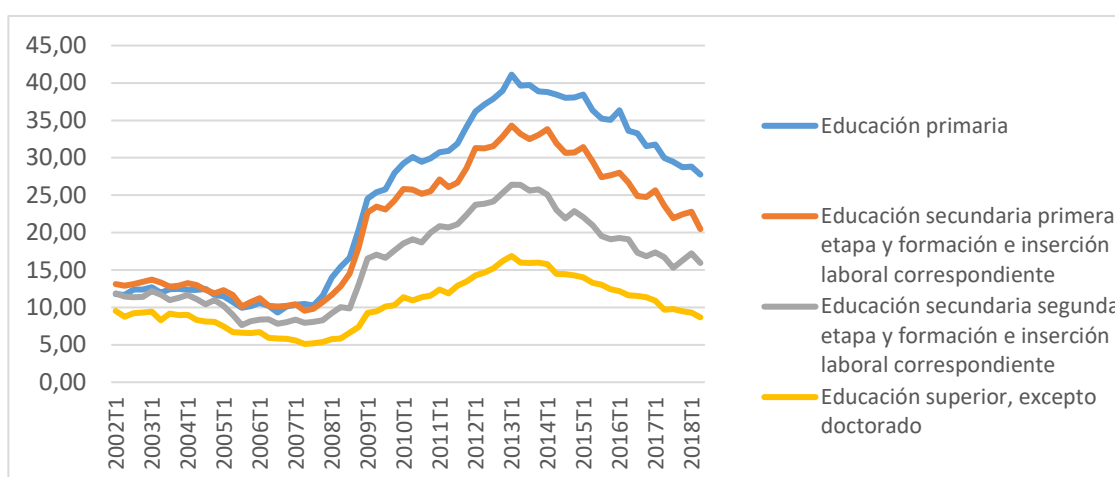


Gráfico 3. Evolución de la tasa de paro por nivel de estudios alcanzados desde 2002T1 hasta 2018T2. Fuente: INE

Por último, y en gran medida consecuencia de las dos relaciones detectadas, la crisis ha tenido un mayor impacto sobre la población inmigrante. Esto se debe a la posición que ocupan estas personas, en los sectores más afectados en la crisis, como la construcción.

Además, también está vinculado con el hecho de que los puestos de trabajo que ocupan tengan bajos requerimientos académicos. Por tanto, se trata de una relación espuria, inducida realmente por una mezcla de los factores anteriores.

2. Fuente de datos

Como hemos comentado anteriormente, los datos con los que trabajamos son de carácter oficial y proceden de la EILU (véase la encuesta completa en el Anexo) [3] procedente del Instituto Nacional Estadística (INE) [4] realizada entre septiembre de 2014 y febrero de 2015. El proceso de recogida de datos fue a través de un cuestionario web y de entrevistas telefónicas. Contamos con 28.905 observaciones y 181 variables cualitativas, pertenecientes a estudiantes graduados en el curso académico de 2009/2010 del conjunto de universidades españolas públicas y privadas. Los datos publicados por el INE han sufrido un proceso previo de depuración y de validación.

A nivel internacional, muchos países realizan desde hace tiempo encuestas que analizan la inserción laboral de los graduados. En cuanto a nivel nacional, algunas universidades realizan este tipo de estudio, pero debido a la variedad de contenido y a las distintas metodologías, hace que no sean comparables los resultados entre sí. También, existen encuestas a nivel regional, ya que algunas CCAA tienen estudios de inserción laboral de los egresados de sus universidades [5].

3. Objetivos

Los objetivos que planteamos en este trabajo son los cinco siguientes:

1. Obtener un perfil general de la población de egresados y por subpoblaciones según su situación laboral.
2. Obtener perfiles de la población de acuerdo a características académicas del individuo.
3. Determinar perfiles laborales de la población según características laborales de los individuos.
4. Detectar relaciones entre los perfiles académicos y laborales.
5. Identificar qué factores influyen a la hora de encontrar empleo y tipo de contrato de este.

4. Metodología

En este apartado haremos una breve descripción de los medios y las técnicas empleadas para conseguir los objetivos del trabajo.

4.1. Recursos informáticos utilizados

En este trabajo vamos a hacer uso de diferentes softwares informáticos. El principal es SAS Base, donde emplearemos los procedimientos correspondientes para cada técnica estadística empleada. También lo usaremos para la creación de nuevos conjuntos de datos y como parte de la preparación de éstos para alcanzar nuestros objetivos.

Los procedimientos *PROC*, ya definidos en SAS Base [6], que usaremos son: *SURVEYSELECT*, *DISTANCE*, *FREQ*, *CLUSTER*, *CORRESP*, *SGPLOT*.

De forma paralela usaremos R, a través del entorno RStudio. En concreto, utilizaremos una función ofrecida por el paquete *klaR* [7], denominada *k-modes* con el fin de aplicarlo a la técnica de clustering. También, con este software realizaremos parte de la preparación de los datos asociada al proceso de utilización de dicha función.

Para el análisis descriptivo hemos optado por SAS Enterprise Miner, ya que este ofrece un gran número de gráficos a los que acceder con facilidad. También ha sido utilizado para la construcción de los árboles de clasificación empleados para el objetivo cinco.

4.2. Técnicas descriptivas

Tanto para la consecución del primero de los objetivos como para interpretar los resultados encontrados a lo largo del trabajo, recurriremos a diversas herramientas de carácter descriptivo.

El desarrollo del primer objetivo consistirá en una descripción de la muestra general, y de las subpoblaciones formadas por los individuos que se encuentran trabajando, en desempleo e inactivos. El análisis será tanto univariante como bivalente.

4.3. Análisis Clúster

Tanto en el objetivo dos como en el tres, los pasos que seguiremos a modo general y que trataremos con más detalle en los siguientes epígrafes, son los siguientes: a. selección de variables, b. Análisis Clúster Jerárquico para determinar el número de grupos, c. Análisis Clúster no Jerárquico. Estamos realizando dos tipos de análisis clúster diferente se debe principalmente a lo siguiente ya que cuando se desconoce el número de grupos que se quiere realizar, es habitual en la literatura realizar un clúster jerárquico de carácter exploratorio que nos indique este número [8], para a continuación, realizar un análisis clúster no jerárquico para el que se necesita saber de antemano el número de grupos. Además, el motivo de hacer este procedimiento es porque el jerárquico se realiza a modo exploratorio, y el no jerárquico ofrece mejores resultados de clasificación.

4.3.1. Selección de variables

El primer paso es la selección de variables que utilizaremos en la determinación de cada perfil. El proceso de selección de variables constará de una serie de pasos, donde el objetivo es obtener un subconjunto de variables con un tamaño reducido que explique al menos el 70% de la variabilidad de los datos en un clúster jerárquico. Para ello:

1. Selección previa de todas las variables relacionadas con aspectos académicos.
2. Pruebas de análisis clúster con diferentes combinaciones de variables, reduciendo progresivamente el tamaño del subconjunto.
3. Selección del subconjunto de variables y clústeres que, en relación con su tamaño, ofrezca mejores resultados.

Una vez que tenemos las variables para cada perfil seleccionadas, continuamos con los siguientes pasos.

4.3.2. Análisis Clúster Jerárquico para determinar el número de clústeres

El análisis clúster jerárquico (ACJ) es una técnica estadística multivariante de clasificación no supervisada que sirve para revelar agrupaciones naturales dentro de un conjunto de datos. Debido a sus características, se trata de una técnica idónea para nuestro objetivo, ya que buscamos clasificar a los individuos en función de una serie de atributos [8]. La idea fundamental de esta técnica es que los individuos dentro de cada grupo sean lo más parecidos entre sí, y los grupos sean lo más diferentes, es decir, heterogeneidad entre grupos. En el ACJ se construye una jerarquía en función de la similitud o distancia de los datos y se obtiene una clasificación para cualquier número de grupos. Por ello, hay que escoger una forma de medir las distancias entre las observaciones y ente grupos de observaciones. Los pasos que sigue esta técnica son:

- I. Se comienza con tantos clústeres como observaciones haya.
- II. Se crea una matriz $n \times n$ que contenga la distancia entre todos los pares de clústeres.
- III. Se agrupan los dos clústeres más próximos, y nos quedamos con un número menor de clústeres.
- IV. Se recalcula la matriz de distancias del punto 2, pero con los clústeres formados en el punto 3.
- V. Se repiten los pasos 3 y 4 hasta que todas las observaciones estén agrupadas en un solo clúster.

Como se puede deducir de los pasos anteriores, la aplicación del análisis clúster jerárquico pasa por concretar:

1. El modo en que se miden las distancias entre elementos.
2. El modo en que se miden las distancias entre clústeres de elementos.
3. Determinación del número de clústeres que responden a una agrupación real.

- **Cálculo de distancias para variables cualitativas**

El ACJ es una técnica propuesta inicialmente para variables cuantitativas, pero que puede ser adaptada a variables cualitativas, como mostraremos a continuación. Cabe recordar que todas nuestras variables son cualitativas.

Los aspectos que difieren con la metodología clásica para variables cuantitativas consisten en la forma de calcular la matriz de distancias inicial (punto II) y en la manera de recalcular la distancia entre clústeres (punto IV).

En cuanto al primer aspecto, los elementos de la matriz de distancias se calculan habitualmente con la distancia euclídea, y en nuestro caso los calcularemos mediante el coeficiente de disimilitud de Jaccard (DJ) [9]. Para calcular este coeficiente se deben obtener previamente variables *dummies* para representar las variables categóricas.

El DJ se define como el número de variables *dummies* coincidentes dividido entre el número total de variables *dummies*. O lo que es lo mismo, uno menos el coeficiente de similitud de Jaccard. En términos matemáticos, el coeficiente de disimilitud entre dos elementos, A y B , viene dado por:

$$DJ(A, B) = 1 - \frac{(A \cap B)}{(A \cup B)}$$

En cuanto al segundo aspecto, al no disponer de las coordenadas de los datos se hace imposible aplicar directamente los métodos de distancias mencionados previamente. Para solventar este problema, en lugar de calcular las distancias entre clústeres directamente, utilizaremos las fórmulas combinatorias propuestas por Lance y Williams [10], dado que solo requieren que se conozcan las distancias entre los clústeres previo a la unión. Por ejemplo, para el cálculo de distancias mediante el método de enlace medio, la fórmula combinatoria que permite calcular la distancia entre un clúster genérico J y uno M resultante de la fusión de otros dos, sería:

$$D_{JM} = \frac{N_K D_{JK} + N_L D_{JL}}{N_M},$$

donde, L y K son los dos clústeres que se fusionan dando como resultado el nuevo clúster M , D_{JM} es la distancia entre los clústeres J y M , y N_K representan el número de observaciones en el clúster K . El resto de fórmulas combinatorias puede consultarse en [11].

- **Distancia entre clústeres**

Con la descripción de los pasos que hay que seguir en el análisis clúster, observamos que, cuando el algoritmo pasa de IV a II, es necesario definir la medida de distancia entre los clústeres. Existen diversos métodos de cálculo de distancias, entre los que probaremos 3, y escogeremos el que ofrezca mejores resultados. Probaremos con los métodos de enlace medio, distancia entre centroides y distancia de Ward o de la mínima varianza [12]. A continuación, los explicamos brevemente:

- Enlace medio: la distancia entre dos clústeres es la distancia media entre observaciones de distintos grupos.
- Distancia entre centroides: la distancia entre dos clústeres es la distancia entre los centroides de cada grupo.
- Distancia de Ward o de la mínima varianza: la distancia utilizada minimiza la varianza dentro de un mismo clúster.

El mejor método será aquel que con un menor número de clúster consiga explicar la máxima variabilidad de los datos posible.

- **Determinar el número adecuado de clústeres**

Una vez obtenida la clasificación para cualquier número de grupos utilizando la metodología que acabamos de explicar, es necesario determinar el número óptimo de clústeres.

Tanto para determinar el número adecuado de clústeres como el mejor método de cálculo de distancias, debemos indicar y definir en qué basaremos nuestra elección. Los indicadores de referencia utilizados en el ACJ son el R^2 , *Pseudo F* y *Pseudo Test de la T* [12]. Todos ellos se basan en los conceptos de variabilidad entre los grupos (E), variabilidad dentro de los grupos (W) y variabilidad total (T) que se obtiene como la suma de los dos anteriores. Recordemos que el objetivo del análisis clúster es obtener unos clústeres homogéneos pero heterogéneos entre sí. Los explicamos a continuación:

- R^2 : Proporción de la variabilidad explicada por los clústeres creados. Buscaremos explicar al menos el 70% de los datos.

$$R^2 = \frac{E}{T}$$

- **Pseudo Test de la T:** si las medias de dos agrupaciones distintas no son significativamente diferentes, entonces esos dos agrupamientos podrían combinarse sin la que la variabilidad dentro del clúster resultante fuese significativamente superior.

Por lo tanto, si al agrupar los clústeres k y l formando m , la dispersión interna de este último clúster (W_m) es mayor que la suma de las dispersiones de los clústeres k y l (W_k y W_l , respectivamente) de forma separada, esa unión no debería llevarse a cabo.

$$Pseudo - T^2 = \frac{W_m - W_k - W_l}{\frac{W_k + W_l}{n_k + n_l - 2}}$$

donde n_k y n_l representan el número de observaciones en los clústeres k y l , respectivamente.

Buscaremos, por tanto, máximos relativos o un incremento considerable, lo que indicará que debemos rechazar ese agrupamiento r y quedarnos en su lugar con $r+1$ clústeres [13].

- **Pseudo F:** este indicador compara la dispersión entre clústeres con la dispersión dentro del clúster. Se busca el valor máximo de este cociente, por lo que nos interesan máximos relativos o variaciones de este estadístico positivas e importantes [13]. A diferencia del anterior, el máximo relativo coincide con el agrupamiento óptimo.

$$Pseudo - F = \frac{\frac{E}{g-1}}{\frac{W}{n-g}},$$

donde n es el número total de observaciones, y g el número de clústeres que se está evaluando.

Por lo tanto, determinaremos el número adecuado de clústeres en función de los valores de estos estadísticos.

4.3.3. Análisis Clúster no Jerárquico: k-modas

Una vez que tengamos el número de clústeres escogido, continuamos con el análisis clúster no jerárquico. Todas las variables con las que estamos trabajando son de tipo categóricas, por lo que no podremos usar el habitual algoritmo *k-medias* [8]. Las opciones para realizar un clúster como este son varias, pero procederemos a realizarlo mediante una generalización del *k-medias*, el algoritmo de agrupamiento el *K-Modas* [14].

Estos dos algoritmos difieren en el método del cálculo del centroide y de distancia. Ambos comienzan eligiendo aleatoriamente tantas observaciones como clústeres se quieran formar (k), y estas se convierten en centroides de cada clúster. Posteriormente, asigna cada una de las $n-k$ observaciones restantes a los k clústeres formados en el paso anterior. Con las nuevas observaciones asignadas en los k grupos, se recalcula el centroide de cada uno, y se verifica que las observaciones están en el clúster cuyo centroide está más cerca. Este proceso se repite hasta que ninguna observación cambie de clúster o se haya alcanzado el número máximo de iteraciones. En el caso del *k-medias* el centroide es el vector de medias de las variables. Mientras que en el *k-modas* el centroide es el vector de modas.

La otra diferencia entre ambos es que, el algoritmo *k-medias* utiliza la distancia euclídea para medir la distancia de una observación a su centroide. En cambio, el algoritmo *k-modas* utiliza el número de variables en las que la categoría de la observación no coincide con la moda del clúster, es decir, el número de categorías discordantes.

Como ya adelantábamos en el apartado de recursos utilizados, utilizaremos la función *k-modes* de la librería **klaR** de R. En ella, basta con indicar el conjunto de datos al que queremos aplicarle el algoritmo de clustering, e indicarle el número de grupos que hemos elegido.

4.3.4. Caracterización de los clústeres

La caracterización de los clústeres es el último paso que llevaremos a cabo dentro para los objetivos dos y tres. Esto consiste en buscar las características más destacadas que definen los clústeres creados. Lo realizaremos sobre el conjunto de variables utilizadas para la formación de los mismos. Para ello, emplearemos técnicas principalmente descriptivas y sus representaciones gráficas, que variarán según su número de categorías:

- Binarias: utilizaremos un mapa de calor, mostraremos en el eje horizontal una de las dos categorías de cada variable, y en el vertical, su representación en porcentaje sobre cada clúster de los creados.
- Más de dos categorías: utilizaremos los gráficos e información de las dimensiones obtenidas de un análisis de correspondencias simples entre la variable clúster y la cualitativa.

Tenemos que aclarar que, esta técnica de correspondencias la hemos empleado para el desarrollo de dos objetivos diferentes a lo largo del trabajo. En primer lugar, como hemos mencionado, la utilizamos para la caracterización de los clústeres obtenidos tanto del perfil académico como del laboral. Pero también para el desarrollo del objetivo cuatro. En este caso, la empleamos para detectar relaciones entre los dos perfiles académico y laboral.

Por lo tanto, en la utilización para la caracterización no evaluamos con demasiado detalle todos los resultados ofrecidos por dicha técnica, pero en el cuarto objetivo sí haremos un análisis exhaustivo.

4.4. Análisis de Correspondencias Simples

El Análisis de Correspondencias Simples (ACS) es una herramienta dentro de las técnicas multivariantes disponibles para variables de tipo categóricas. El objetivo es presentar cada una de las modalidades posibles de cada una de las variables en cuestión en un plano donde la posición relativa de los puntos refleje el grado de asociación entre cada una de las modalidades representadas. Se utiliza con el fin de encontrar relaciones entre las categorías de dos variables con una reducción de la dimensionalidad [15].

En el caso del objetivo cuatro, las variables analizadas se corresponden con los dos grupos de clústeres formados para los dos perfiles, académico y laboral. Los clústeres que conforman cada perfil serán las categorías/modalidades de cada variable.

Esta técnica la llevaremos a cabo en SAS Base, el cual tiene el procedimiento *proc corresp*, y que proporciona todo lo que nos interesa analizar. Los pasos habituales a seguir en el ACS los enumeramos y explicamos a continuación [12].

- **Estudio de las proporciones sobre el total de la población de cada modalidad**

Una de las salidas del procedimiento es una tabla de contingencia con las modalidades de las variables en cuestión en la que aparece, expresado en porcentaje, la distribución de las categorías de cada variable. Debemos prestar atención a aquellas que tengan una presencia inferior al 5%. El motivo por el cual se establece el 5% es debido al funcionamiento del contraste de independencia que explicaremos a continuación. En caso de existir alguna modalidad con baja representación, nos fijaremos si está bien representada sobre los ejes, y de ser así, continuaremos con el análisis. En caso contrario, estudiaremos la posibilidad de agrupar dichas categorías con otras, o bien considerarla como suplementaria.

- **Contraste de independencia de variables, y estudio de las aportaciones al Chi-cuadrado**

Contraste de independencia

La condición para que tenga interés el ACS es que las variables no sean independientes entre sí. Consideraremos que dos variables son independientes cuando el valor de una variable no influye sobre la distribución de la otra. Es decir, la probabilidad de que una variable tome un valor no dependerá del valor que tome la otra.

Para contrastar si existe relación entre ambas variables se llevará a cabo un contraste de hipótesis sobre la independencia de variables. El estadístico habitualmente utilizado es el de Chi-Cuadrado de Pearson.

En este contraste se presupone la independencia de variables en la hipótesis nula y permite evaluar las desviaciones de una muestra respecto al valor teórico previsto.

H₀: Las variables son independientes

H₁: Las variables no son independientes

El estadístico asociado al test se define como:

$$\sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - e_{ij})^2}{e_{ij}} = \chi^2_{(r-1) \cdot (c-1)}$$

Donde,

n_{ij} : frecuencia observada del cruce de categorías i y j .

e_{ij} : frecuencia esperada bajo H_0 del cruce de categorías i y j .

Y se distribuye según una distribución χ^2 con $(r-1) \cdot (c-1)$ grados de libertad. Donde r es el número de categorías de la variable considerada en filas, y c de columnas.

Buscamos que la diferencia entre lo observado y lo esperado sea suficientemente grande como para rechazar la hipótesis nula. Es decir, valores grandes del estadístico Chi-cuadrado son indicativos de una fuerte relación entre las variables estudiadas.

El valor del estadístico y la probabilidad de comportamiento de la muestra si H_0 fuera cierta vienen dados en una de las salidas ofrecida por el procedimiento *proc corresp*. Si el p-valor es suficientemente pequeño podremos rechazar la hipótesis nula, y continuar con el análisis.

Aportaciones al chi-cuadrado

Dado que el estadístico anterior se obtiene como la suma de las aportaciones al Chi-Cuadrado de los cruces de las modalidades, un aspecto interesante sería saber cuáles de esos cruces contribuyen más el valor final del estadístico. Para ello, mostraremos un mapa de calor donde la intensidad del color viene dada por la aportación al valor final del estadístico Chi-Cuadrado del cruce de modalidades. Para saber el tipo de la relación (directa o inversa) recurriremos a la tabla *Observed Minus Expected Values* donde se muestra la diferencia entre las frecuencias observadas y las esperadas.

Por ejemplo, un valor negativo en esa tabla señalará que hay menos observaciones en ese cruce de categorías de las que debería haber bajo la hipótesis de independencia, y, por tanto, esas dos categorías estarán relacionadas inversamente, es decir, presentarán disociación.

- **Análisis con las observaciones y modalidades elegidas**
 - **Determinar el número de dimensiones que hay que retener**

La elección del número de dimensiones que hay que retener la basaremos en dos criterios frecuentemente utilizados, ambos se basan en la inercia. Esta es una medida de dispersión de las distribuciones condicionadas filas y columnas, respectivamente, de la tabla de contingencia, que tiene en cuenta las dos variables al mismo tiempo, y que toma valores bajos cuando las variables son independientes, y valores altos en caso contrario. Esta información de nuevo está disponible entre las salidas del procedimiento SAS. Consideraremos como criterio principal escoger tantas dimensiones como sean necesarias para explicar al menos el 90% de la inercia. Adicionalmente, utilizaremos la regla conocida como *average rule*, que consiste en escoger aquellas dimensiones cuya inercia sea superior a la media que se define como:

$$\frac{Inercia}{\min\{r - 1, c - 1\}}$$

Donde,

r : número de categorías de la variable considerada en filas

c : número de categorías de la variable considerada en columnas

- o **Analizar la elección del número de dimensiones mediante un examen de la calidad de la representación de la modalidad sobre el eje**

Las dimensiones escogidas en el punto anterior deben representar las modalidades correctamente sobre los ejes. Esto lo comprobaremos mediante el análisis de las calidades de las modalidades. Esta calidad mide la proporción de la inercia de cada modalidad que es explicada por los ejes retenidos. Supondremos que las modalidades están bien representadas si tienen al menos una calidad del 70%.

- o **Interpretar las dimensiones a través de las correlaciones entre las dimensiones y las modalidades (cosenos al cuadrado), y de la contribución parcial de cada modalidad**

En este punto trataremos de definir cada dimensión, para lo que puede ser útil conocer qué modalidades son las que más contribuyen a la inercia de las mismas.

Para encontrar la correlación entre los ejes y las modalidades nos fijaremos en los cosenos al cuadrado que son la correlación entre la modalidad y el eje. Los valores son siempre positivos por lo que también nos fijaremos en el signo de las coordenadas, para identificar el tipo de relación. Una vez identificada la proporción de cada categoría explicada por cada dimensión, trataremos de buscar cuales son las modalidades que mejor explican las dimensiones, de nuevo para cada variable. Para ello, nos fijaremos en la contribución parcial a la inercia de cada categoría para cada dimensión.

- o **Representación gráfica y descubrimiento de relaciones**

Representaremos gráficamente las categorías de las dos variables sobre los ejes escogidos. A continuación, explicaremos las relaciones existentes entre las modalidades de las dos variables teniendo en cuenta la proximidad entre modalidades de la misma variable y los ángulos formados entre categorías de ambas variables. Cabe destacar que si las modalidades se encuentran en cuadrantes opuestos presentarán disociación. Por el contrario, si el ángulo formado entre dos modalidades (teniendo como referencia el eje de coordenadas) es pequeño, estas estarán relacionadas.

4.5. Árboles de clasificación

Los árboles de clasificación son una técnica de clasificación supervisada. Constituyen una herramienta útil para la segmentación de datos a partir de una serie de reglas simples, que se van aplicando de forma jerárquica y secuencial. Partiendo del conjunto de datos completo, estas reglas dan lugar a subconjuntos de la población, que se denominan nodos, y que tienen la propiedad de ser más homogéneos con respecto a la variable de interés que el nodo anterior.

Entre las principales ventajas de los árboles se encuentran:

- La sencillez de su interpretación.
- La falta de suposiciones teóricas sobre los datos.
- Permiten encontrar interacciones y reglas difíciles de encontrar con otros métodos.
- Aportan medidas de importancia de las variables.
- Funcionan bien en presencia de datos atípicos y ausentes.

Las fases de la construcción de un árbol son:

1. Se comienza con un nodo inicial (llamado raíz), que contiene todas las observaciones del conjunto de datos.
2. De entre todas las variables de entrada disponibles, se selecciona aquella que divida las observaciones del nodo precedente en dos partes más homogéneas.
3. De este nodo *padre* saldrán ahora dos nodos *hijos*. En cada uno de estos nodos se vuelve a repetir el proceso de seleccionar una variable y un punto de corte para dividir la muestra en dos partes más homogéneas.
4. El proceso termina cuando ocurra algo de lo siguiente:
 - a. Todas las observaciones de un mismo nodo sean del mismo tipo.
 - b. Ninguna de las variables disponibles produzca una mejora suficiente.
 - c. Todas las posibles divisiones den lugar a *hijos* con pocas observaciones.
 - d. La profundidad del árbol (distancia, en número de nodos, entre el nodo raíz y el nodo más alejado) supere el umbral predefinido.

La forma de seleccionar las variables de segmentación y su punto de corte se basa en distintos criterios como pueden ser: estadístico Chi-Cuadrado (dado por la fórmula vista en el apartado anterior), índice de Gini o Entropía. Para más información acerca de estos criterios ver [16]. Además, existen otros parámetros que determinan la forma del árbol final, y, por tanto, la bondad del ajuste. Entre ellos, destacamos la profundidad del árbol, el número de observaciones mínimo en cada nodo, el nivel de significación en el caso del Chi-Cuadrado, etc.

Como es habitual en minería de datos, evaluaremos el árbol a partir de un subconjunto de datos de la muestra inicial, denominado *prueba*, que no ha sido utilizado para la construcción del modelo, de manera que podemos obtener una medida realista e insesgada. Los datos restantes conforman el conjunto de datos con los que se construye el modelo, denominado *entrenamiento*. De cara a reducir el efecto de la aleatoriedad lo máximo posible, el proceso de partición de los datos *entrenamiento-prueba* se repetirá con distintas semillas para evaluar el comportamiento del modelo, tanto en media como en varianza. En cada una de esas repeticiones se obtendrá la medida de evaluación.

En nuestro caso, la medida que vamos a utilizar tanto para evaluar el modelo como para escoger el mejor de entre los que probaremos, es la *tasa de clasificación errónea* que, como su propio nombre indica, cuantifica el porcentaje de observaciones mal clasificadas. Por lo tanto, buscaremos que este valor sea lo menor posible.

Tal y como hemos comentado, los árboles permiten obtener una medida de la importancia de las variables. Para ello, se calcula la mejora que provoca la introducción de una nueva variable en el modelo. A partir de estas cantidades, genera un ranking de

las variables involucradas en el modelo, permitiendo saber cuáles son las más influyentes.

5. Perfil general de la muestra y de la subpoblación de egresados

La población de la *Encuesta de inserción laboral de titulados universitarios* son titulados en universidades españolas tanto públicas como privadas en el curso académico 2009/2010. Los detalles de la muestra se recogen en el capítulo dos.

Las variables que se consideran recogen información acerca de la situación laboral de los individuos y del proceso de acceso al mercado laboral. Según la metodología de la encuesta, se agrupan en las siguientes categorías:

1. Información de identificación de los titulados: sexo, edad, nacionalidad
2. Características de los titulados universitarios del curso 2009-10: titulación, tipo de universidad, rama de conocimiento, etc.
3. Lugar de residencia en 2014.
4. Situación laboral en 2014.
5. Trabajo actual: situación profesional, tipo de jornada, ocupación desempeñada, tamaño de la empresa, etc.
6. Información relacionada con la subpoblación inactiva.
7. Primer empleo tras acabar la carrera: tiempo transcurrido hasta el primer empleo, situación profesional, tipo de jornada, etc.
8. Historial laboral: número de empleadores, tiempo total trabajado, etc.
9. Trabajo remunerado mientras estudiaban.
10. Egresados que no han trabajado tras titularse: búsqueda de empleo, tiempo buscando empleo, forma de la búsqueda, etc.
11. Formación y aprendizaje: prácticas curriculares y extracurriculares, becas, conocimiento de idiomas, habilidades informáticas, etc.
12. Movilidad.
13. Seguridad social: bases de cotización de los afiliados por cuenta ajena, etc.
14. Cursos de formación del Servicio Público de Empleo Estatal (SEPE).

Para conocer los valores que toman las variables anteriores y poder, por tanto, conocer el perfil de la muestra completa, realizamos un análisis descriptivo univariante y bivalente, del que mostraremos las variables más informativas.

• Análisis descriptivo

En primer lugar, nos fijaremos en las variables recogidas en la categoría “información de identificación de los titulados”. Nos encontramos ante una muestra cuya nacionalidad es en un 99,30% española, y el 60% género femenino. Como se puede observar en el Gráfico 4, el tramo de edad mayoritario es *menores de 30 años*, con una representación del 58%.

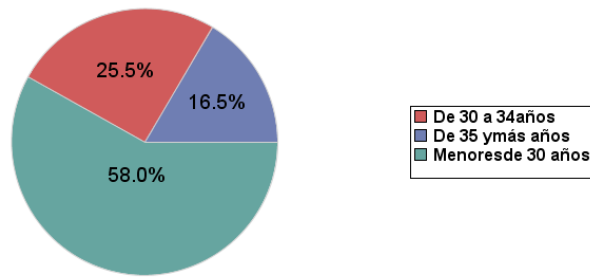


Gráfico 4. Distribución de la edad de los egresados

A continuación, mostraremos la distribución de la situación laboral de la muestra. En el Gráfico 5 vemos que el 76,4% de los individuos se encuentra trabajando, y el 18% desempleado, y el 5,6% son inactivos.

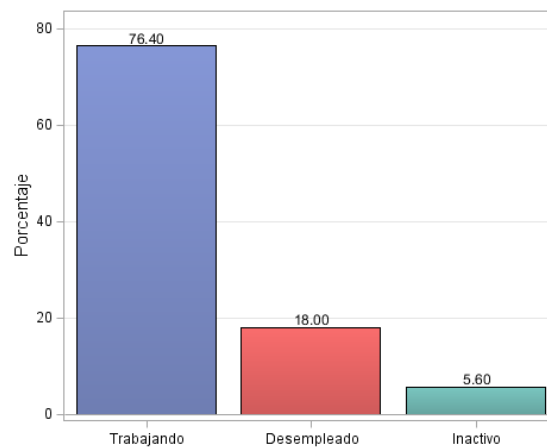


Gráfico 5. Distribución de la situación laboral de los egresados

En cuanto al ámbito académico, en el Gráfico 6 vemos la distribución del tipo de universidad en la que han estudiado los individuos, pública o privada. Casi un 85% de la población muestreada ha estudiado en una universidad pública.

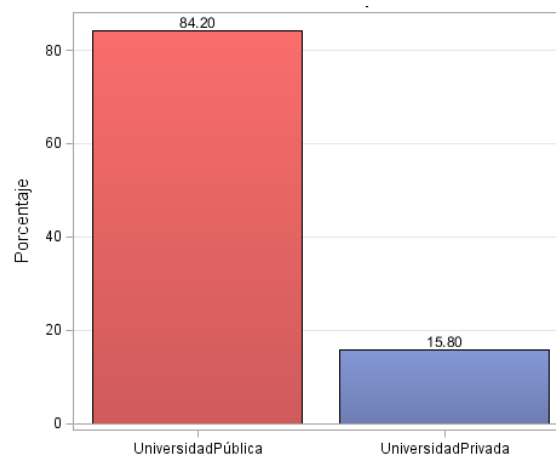


Gráfico 6. Distribución del tipo de universidad en la que estudiaron los egresados

La distribución de las ramas de estudio la vemos en el Gráfico 7. En ella vemos que la rama con más representación es la de ciencias sociales y jurídicas, y a continuación la rama de ingeniería y arquitectura.

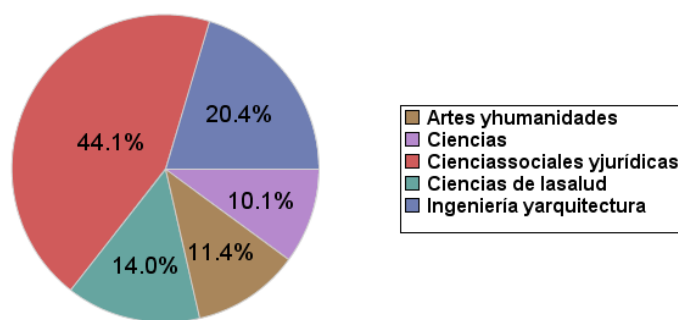


Gráfico 7. Distribución de la rama de estudio de los egresados

➤ Perfiles de la subpoblación según su situación laboral

A continuación, describiremos a la población según se encuentren trabajando, desempleados o inactivos. Para ello, emplearemos las variables utilizadas anteriormente.

En la Tabla 1, mostramos la distribución de la rama de estudio para cada uno de los 3 grupos junto con el perfil general. En todos los grupos, la rama mayoritaria es la de ciencias sociales y jurídicas, aunque tiene aproximadamente un 4% de representación más entre los desempleados con respecto a la población general, y los que se encuentran trabajando.

Otro aspecto llamativo es la rama de ciencias de la salud. Se aprecia una fuerte disminución en el grupo de los desempleados llegando a un 7,2%, siendo la mitad del 14% que suponía esta rama entre los egresados.

La rama de artes y humanidades destaca por la fuerte presencia entre los inactivos 19,64%, en comparación con el 11,4% que hay entre todos los titulados.

Por otro lado, la rama de ciencias tiene menos presencia entre los desempleados, con un 8,3% de representación; mientras que para los inactivos este porcentaje llega hasta el 14,2%.

Por último, la rama de ingeniería y arquitectura tiene un 10,7% de presencia entre los inactivos, lo cual destaca porque es la mitad del peso de dicha rama entre el conjunto de todos los egresados.

Tabla 1. Distribución de la rama de estudio para cada situación laboral

Rama	Trabajando	Desempleado	Inactivo	General
Ciencias sociales y jurídicas	43,06%	48,33%	44,64%	44,10%
Ciencias de la salud	15,84%	7,22%	10,71%	14,00%
Artes y humanidades	10,21%	13,89%	19,64%	11,40%
Ciencias	10,21%	8,33%	14,29%	10,10%
Ingeniería y arquitectura	20,68%	22,22%	10,7%	20,40%

En cuanto a la distribución del sexo según su situación laboral, vemos en el Gráfico 8 que existe un mayor porcentaje de mujeres entre los desempleados. Hay una proporción mayor entre los desempleados comparado con el 60% que existe en la muestra general, y en cuanto a los inactivos, esta proporción es menor.

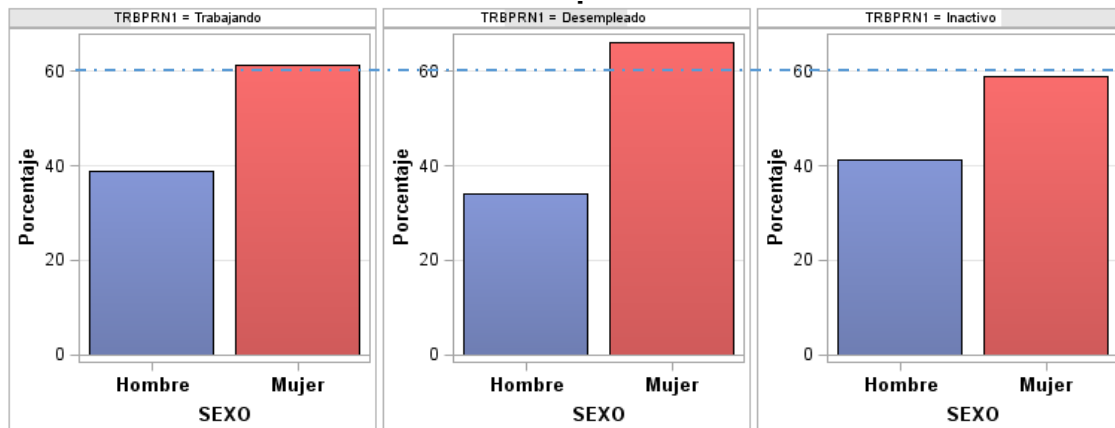


Gráfico 8. Distribución del sexo por situación laboral

En el Gráfico 9 vemos la distribución del tipo de universidad para cada grupo. En el caso de los desempleados, la representación de la universidad privada es menor que en los que se encuentran trabajando o inactivos.

Tras este análisis exploratorio de los datos podemos decir que:

- La presencia de mujeres en el grupo de desempleados es mayor que en la población general.
- El porcentaje de estudiantes de universidad pública se incrementa entre los desempleados.
- La distribución de las ramas de estudio varía según la situación laboral. La rama de ingeniería y arquitectura duplica su presencia entre los desempleados con respecto a los inactivos. Y la rama de artes y humanidades presenta un 9% más entre los inactivos con respecto a los que están trabajando.
- La rama de ciencias de la salud tiene una menor presencia en los individuos desempleados e inactivos, con respecto a los que están trabajando.

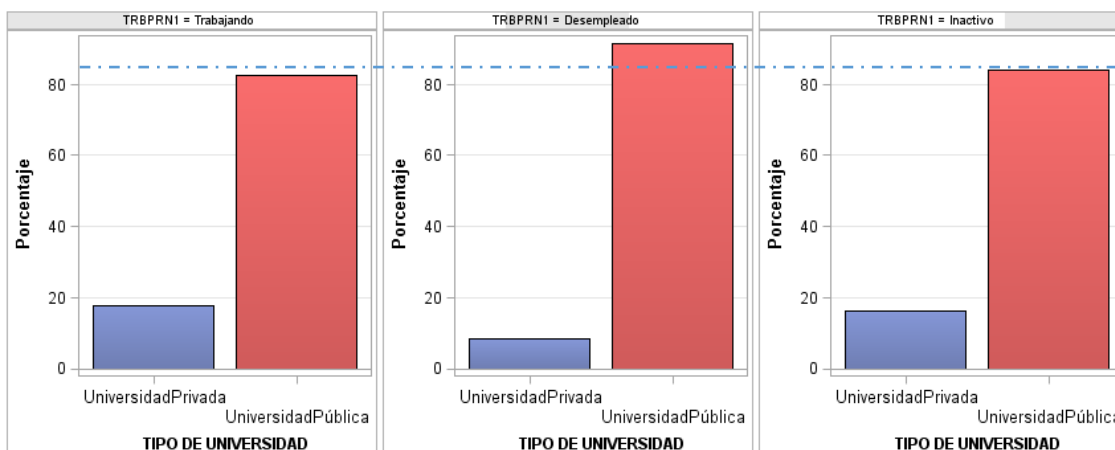


Gráfico 9. Distribución del tipo de universidad por situación laboral

A continuación, enfrentamos la rama de la titulación frente al tipo de universidad. A su vez, haremos distinción entre los que están trabajando y los desempleados. Para los que están trabajando mostraremos el Gráfico 10 de celosía de barras.

En primer lugar, vemos la distribución de la rama por tipo de universidad de los recién egresados que se encuentran trabajando actualmente. En todas las ramas, predomina los estudios en universidad pública a excepción de la ingeniería y arquitectura, y ciencias de la salud, cuya representación es mayor en la universidad privada. También podemos apreciar como la rama de artes y humanidades tiene una proporción mayor entre las universidades públicas que en las privadas.

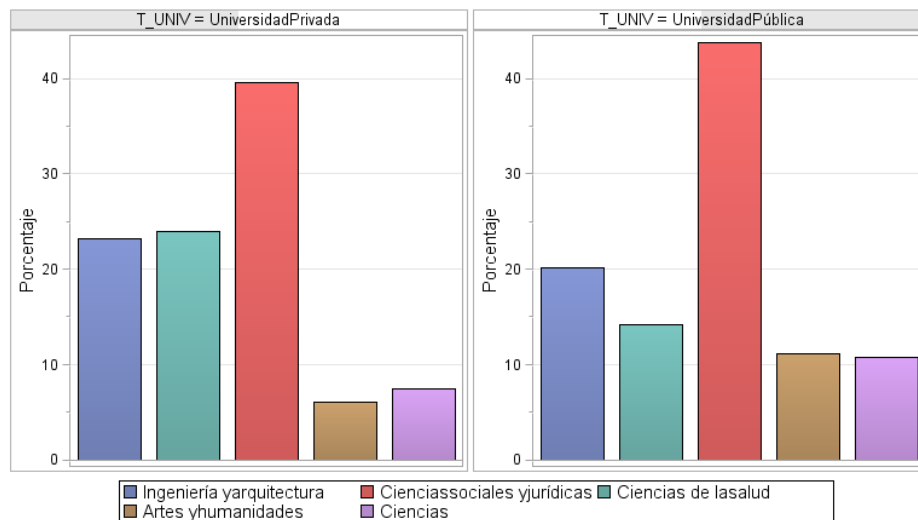


Gráfico 10. Distribución de la rama de estudio por tipo de universidad de egresados que se encuentran trabajando

En el caso de los desempleados, lo primero que llama nuestra atención en el Gráfico 11 es que no existe representación de la rama de ciencias de la salud en universidad privada. Por lo que de todo nuestro conjunto de datos ninguno de los estudiantes de ciencias de salud que estudiaron en una universidad privada se encuentra en desempleo.

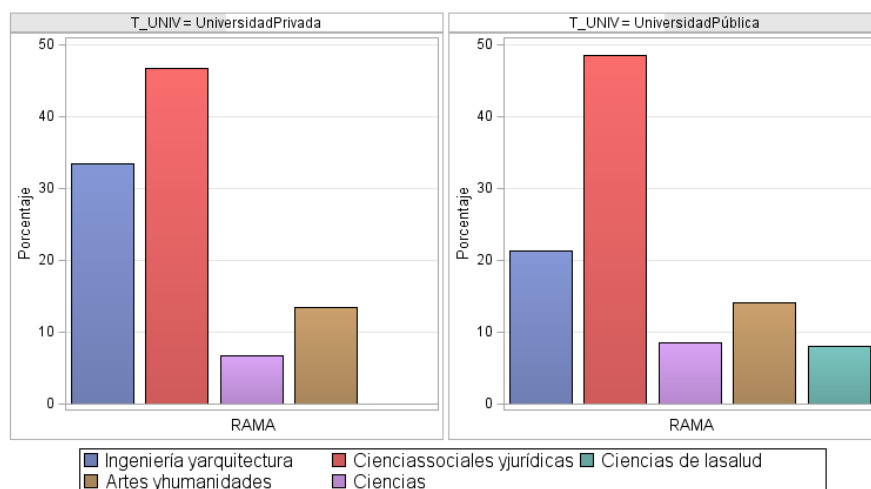


Gráfico 11. Distribución de la rama de estudio por tipo de universidad de egresados que se encuentran desempleados

Hemos analizado la influencia de la posesión de una acreditación oficial de idiomas en la situación laboral del egresado. Existen dos variables de dos certificaciones diferentes:

- Título de la Escuela Oficial de Idiomas
- Título en lengua extranjera reconocido oficialmente (PET, FCE, CAE, CPE, DELF, DALF, ...etc)

Así que las hemos agrupado en una nueva variable (llamada *IDIOMA_OFICIAL*) que recoge la información de si el individuo tiene idioma reconocido oficialmente o no.

Por medio del componente de celosías, en SAS EM, representamos en el Gráfico 12, donde a la izquierda los individuos tienen el idioma acreditado oficialmente, y a la derecha lo contrario.

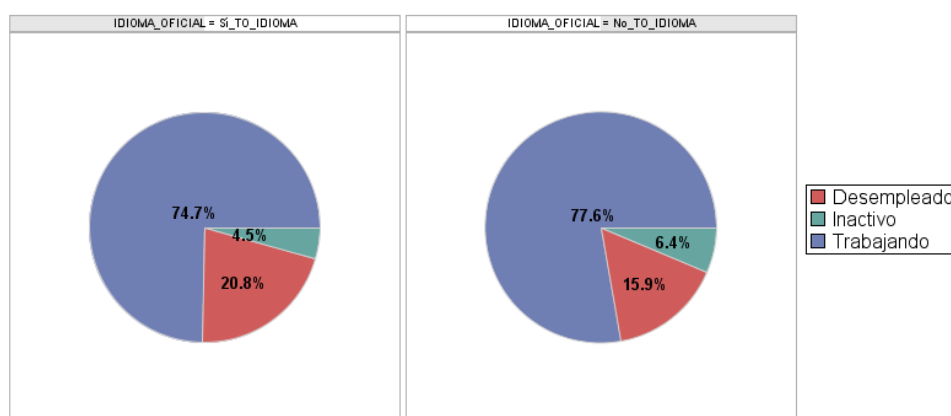


Gráfico 12. Distribución del idioma acreditado oficialmente por situación laboral

Podemos observar que no existen diferencias obvias, lo cual no quiere decir que esto no influya, sino que puede que en presencia de otros factores si lo haga, pero individualmente no.

➤ Contraste de independencia entre variables

Para contrastar la significatividad de las diferencias observadas en los gráficos anteriores, a continuación realizaremos una serie de contraste de independencia Chi-Cuadrado para determinar si existe relación entre la situación laboral y las siguientes variables: sexo, rama, idioma oficial y tipo de universidad. Los resultados se encuentran en la Tabla 2.

Tabla 2. Resultados del contraste de independencia entre variables y situación laboral

Variables	Grados de libertad	Chi-Cuadrado	p-valor
Sexo	2	36,7640	<0.0001
Rama	8	399,7424	<0.0001
Idioma oficial	2	1,2451	0.5366
Tipo de universidad	2	93,2440	<0.0001

Como podemos observar en la tabla, todas las variables menos el *Idioma Oficial* presentan diferencias significativas con respecto a la situación laboral.

➤ Análisis adicional para los individuos que están trabajando

Otro aspecto interesante, de acuerdo a los objetivos que planteamos en este trabajo, es conocer la distribución de los individuos en cuanto a la situación contractual de los que están trabajando. La variable que recoge esta información cuenta con 6 categorías: asalariado con contrato temporal, asalariado con contrato permanente, empresario o trabajador independiente, ayuda familiar y en prácticas, formación o becario. Hemos decidido representar esta variable para hombres y mujeres en el Gráfico 13.

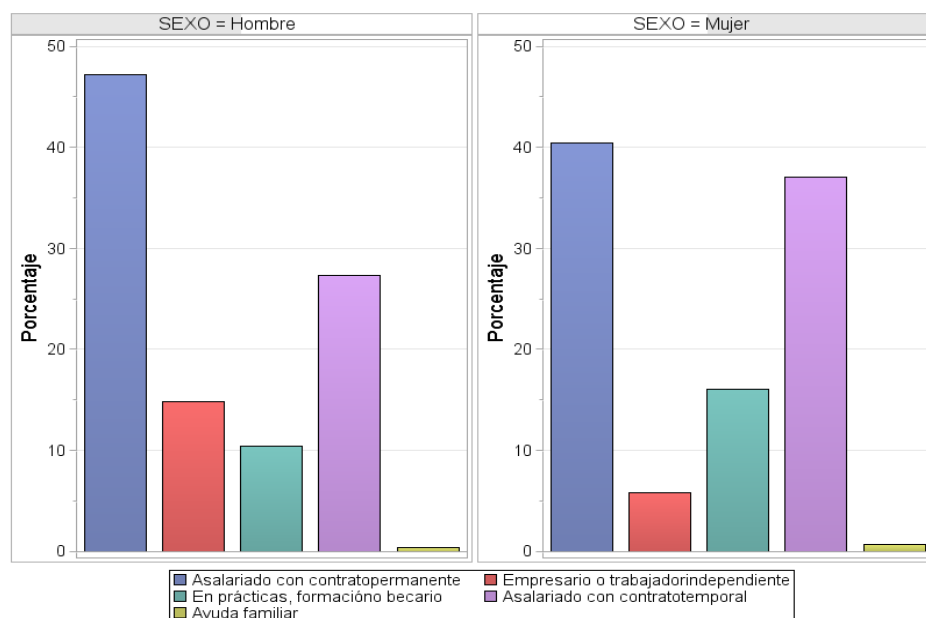


Gráfico 13. Distribución del tipo de contrato por sexo

En los asalariados con contrato permanente y temporales, y empresarios o trabajadores independientes es donde encontramos mayor diferencia entre ambos géneros. En ambos casos, la proporción es mayor para los hombres que para las mujeres. En cambio, la proporción de contrato temporal es mayor para las mujeres que para los hombres. Ocurre lo mismo para las prácticas, formación y becario, y ayuda familiar.

A continuación, en el Gráfico 14 vemos la distribución de la base de cotización a la seguridad social en 2014 para cada género. Esta variable consta de cinco categorías que corresponden al nivel de cotización a la seguridad social de los empleados por cuenta ajena, y se divide en cinco rangos denominados quintiles: primero, segundo, tercero, cuarto y quinto. Los individuos que sean empresarios independientes, es decir, autónomos no pertenecerán a ninguno de estos quintiles, y se encontrarán en la categoría denominada en el conjunto de datos *no aplicable*.

En los quintiles bajos de cotización, no vemos grandes diferencias entre hombres y mujeres. Es en el quinto quintil donde vemos uno de los aspectos más relevantes. Y es que, en el caso de los hombres, este quintil representa el 15,22% de la población. Mientras que, para las mujeres, es prácticamente la mitad, un 8,24%. Lo mismo ocurre, pero al revés, para el primer y segundo quintil. En este caso, la proporción de esta categoría entre las mujeres es del 13,89% y 14,86%, respectivamente. Mientras que para los hombres es un 8,14% para el primer quintil y 7,87%, para el segundo. En ambos

casos, la categoría *no aplicable*, que se corresponde con los autónomos, es la categoría mayoritaria.

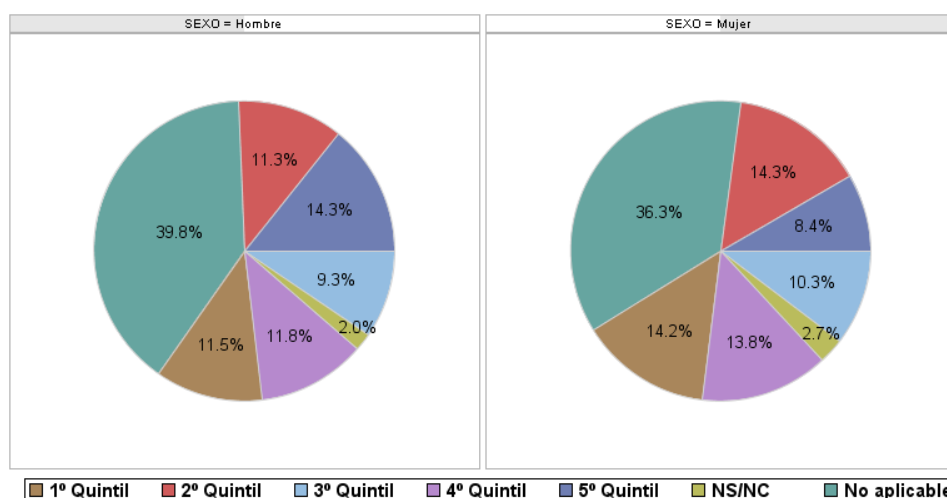


Gráfico 14. Distribución por sexo de la base de cotización a la Seguridad Social en 2014

Una vez que hemos descrito la muestra con la que trabajaremos, el siguiente paso es comenzar a desarrollar los objetivos propuestos. Procederemos tal y como hemos descrito en la metodología.

6. Clasificación de los grupos de perfiles académicos

La información con la que contamos y con la que comenzaremos este análisis está recogida en las siguientes variables agrupadas según las categorías propuestas en la metodología de la encuesta [5].

1. Información de identificación de los titulados:
 - Sexo
 - Edad
2. Características de los titulados universitarios del curso 2009-10
 - Titulación realizada
 - Tipo de universidad
 - Rama y ámbito de conocimiento
 - Discapacidad
3. Formación y aprendizaje
 - Prácticas curriculares y extracurriculares
 - Becas
 - Estudios universitarios realizados en el extranjero
 - Otra formación realizada (de grado medio y grado superior)
 - Formación sanitaria especializada
 - Realización de otros estudios oficiales en el momento de la entrevista
 - Conocimiento de idiomas
 - Habilidades informáticas

6.1. Selección de variables

- **Selección de una muestra estratificada de tipo instrumental**

Al comenzar esta selección de variables nos dimos cuenta que nuestros recursos informáticos no podían realizar el procedimiento *proc cluster* con las 28.905 observaciones que partíamos.

Tras una búsqueda bibliográfica sobre este aspecto, descubrimos en el artículo de Zhexue Huang [14] que el coste cuadrático computacional que implica realizar el clúster jerárquico con la matriz de distancias, es muy elevado y no es posible hacerlo con conjuntos de datos muy grandes.

Es por ello que hemos decidido realizar un muestreo estratificado de tipo instrumental por la variable relacionada con la información de la situación laboral del individuo, para asegurarnos que los desempleados e inactivos queden representados en la muestra, puesto que solo representan el 19,01% y 6,7%, respectivamente, de los datos iniciales y nos interesan en relación con su perfil académico.

En la Tabla 3 vemos los estratos, el tamaño de la muestra y el porcentaje de afijación (afijación proporcional).

Tabla 3. Afijación proporcional de la muestra estratificada para los perfiles académicos

Estratos	Tamaño	% (afijación)
Empleados	21484	74,3
Desempleados	5494	19,01
Inactivos	1927	6,7

La sentencia empleada para ello en SAS Base es la siguiente, con la que hemos seleccionado una muestra de 20.000 observaciones, que tras ciertas pruebas computacionales, es lo máximo que nos ha permitido llevar a cabo el clúster jerárquico.

```
PROC SORT DATA=trabajo.todos  
OUT=trabajo.todos1 ;  
BY TRBPRN1 ;  
RUN;  
PROC SURVEYSELECT DATA=trabajo.todos1  
OUT=trabajo.todostra N=20000 seed=12345 ;  
STRATA TRBPRN1 / alloc=prop;  
RUN;
```

Una vez tenemos un conjunto de datos con el que poder realizar el ACJ, procedemos a seleccionar las variables. Entre las variables disponibles asociadas a la titulación del individuo, hemos decidido escoger la rama de conocimiento, dado el número elevado de categorías con las que cuenta la titulación y el ámbito. Hemos considerado tres variables fijas de acuerdo con nuestro objetivo, y son el sexo del individuo, la rama de estudios y el tipo de universidad.

La Tabla 3 muestra un resumen de los mejores resultados obtenidos de las distintas combinaciones de variables de las pruebas realizadas. En la primera columna indicamos el número de variables utilizadas. En la segunda, vemos el número de clústeres necesarios para explicar la variabilidad recogida (medida a partir del estadístico R^2

explicado en la metodología) en la tercera columna. Más detalle del proceso de estas pruebas, pueden consultarse en el anexo.

Tabla 4. Resultado de las pruebas de selección de variables para perfiles académicos

Pruebas	Número de clústeres en la mejor combinación de variables	Variabilidad explicada
Pruebas con 10 variables	30	70,4%
Pruebas con 9 variables	26	72,3%
Pruebas con 8 variables	14	72,1%
Pruebas con 7 variables	15	74,3%
Pruebas con 6 variables	10	74,4%

Finalmente, escogemos uno de los resultados de las pruebas con 6 variables, que con 10 clústeres consiguen explicar el 74,4% de la variabilidad. Las variables escogidas para la creación de los conglomerados son las que mostramos en la Tabla 4.

Tabla 5. Variables seleccionadas para obtener los perfiles académicos

Variable	Descripción	Categorías
SEXO	Sexo	Hombre Mujer
RAMA	Rama de conocimiento de la titulación	Ciencias sociales y jurídicas Ciencias de la salud Ciencias Artes y humanidades Ingeniería y arquitectura
T_UNIV	Tipo de Universidad	Universidad Pública Universidad Privada
EST_B1	Disfrutó de alguna beca durante sus estudios	Sí No
EST2_MA	Ha completado un máster en una universidad española	Sí No
IDIOMA_OFICIAL	Posee título oficial de idiomas	Sí No

6.2. Análisis clúster jerárquico

Una vez hemos determinado el conjunto de variables podemos proceder a determinar el número de grupos óptimo con la muestra estratificada instrumental creada anteriormente.

- **Matriz de distancias**

Para el cálculo de la matriz de distancias emplearemos la siguiente sentencia SAS. En ella debemos indicar que el método escogido es *djaccard* y las variables seleccionadas.

```
proc distance data=trabajo.todostra1 method=djaccard absent=0 out=trabajo.distjac;
var anominal(SEXO RAMA T_UNIV EST_B1 EST2_MA IDIOMA_OFICIAL);
id IDEN;
run;
```

- **Determinar el número de grupos**

El conjunto de distancias calculado en el paso anterior será la entrada para el procedimiento *proc cluster* que llevaremos a cabo a continuación.

Recordemos que buscamos máximos relativos en el estadístico *Pseudo F*, y valores pequeños del *Pseudo T Squared* seguidos de un máximo relativo. El número de agrupaciones mínimo que empezaremos a valorar será a partir de un R^2 superior al 70%. Además, seremos comedidos en cuanto al número se refiere puesto que esto podría dificultar su visualización en los apartados anteriores.

- **Elección del mejor método de cálculo de distancias entre clústeres**

La Tabla 6 recoge la elección de grupos con cada método, y el porcentaje de variabilidad explicada con cada uno. Para más detalle de estos resultados, puede consultarse el anexo.

Tabla 6. Resultados de la elección del método de cálculo de distancias entre clústeres de los perfiles académicos

Métodos de cálculo de distancias	Agrupaciones	Variabilidad explicada
Enlace medio	16	73%
Distancia entre centroides	14	70,4%
Distancia de Ward	10	74,4%

Tras realizar dichas pruebas y con la información de la tabla anterior, concluimos que fue el de Ward el que mejores resultados ofreció, ya que un con número bastante menor de clústeres conseguimos explicar prácticamente lo mismo que los demás métodos con más grupos.

```
proc cluster data=trabajo.distjac(type=distance) method=WARD pseudo outtree=ttree
RSQUARE print=30;
var O4--O28904;
id IDEN;
run;
```

A continuación, desarrollamos los motivos por los que escogimos ese número de grupos con el método de Ward. Mostraremos la tabla que recoge la información de los estadísticos y su gráfica asociada. Por brevedad, solo mostraremos 10 de los 30 grupos representados en SAS Base. El Gráfico 15 que recoge el valor de los estadísticos en cada agrupación, vemos que en el caso del *Pseudo F* las variaciones no son tan acentuadas como en el caso del *Pseudo T Squared*.

Siguiendo las explicaciones de la metodología, en el clúster 9 observamos un máximo relativo en el *Pseudo T Squared*, y a continuación un fuerte descenso del mismo. Esto mismo lo podemos comprobar analizando la Tabla 7. Por todo ello, el número de clústeres escogidos será 10. Además, conseguimos explicar con ellos el 74,5% de la variabilidad.

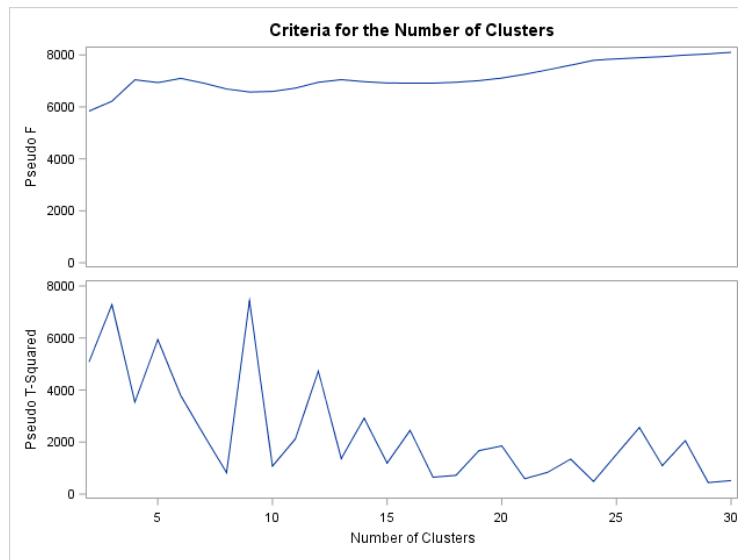


Gráfico 15. Pseudo F y Pseudo T para la selección del número de clústeres de los perfiles académicos

Tabla 7. Estadísticos para la selección del número de clústeres de los perfiles académicos

Cluster History							
Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square	Pseudo F Statistic	Pseudo t-Squared
15	CL19	CL39	2032	0.0095	.829	6920	1193
14	CL40	CL43	1667	0.0097	.819	6968	2918
13	CL22	CL25	2862	0.0104	.809	7048	1366
12	CL35	CL31	2730	0.0161	.793	6949	4730
11	CL20	CL23	2434	0.0219	.771	6723	2124
10	CL21	CL17	1868	0.0228	.748	6595	1076
9	CL26	CL28	4092	0.0235	.725	6575	7428
8	CL10	CL18	2763	0.0239	.701	6688	823
7	CL15	CL16	3452	0.0259	.675	6915	2288
6	CL12	CL14	4397	0.0351	.640	7100	3788
5	CL9	CL13	6954	0.0586	.581	6935	5942
4	CL7	CL11	5886	0.0674	.514	7043	3540
3	CL5	CL6	11351	0.1303	.383	6219	7285
2	CL8	CL3	14114	0.1574	.226	5840	5082
1	CL4	CL2	20000	0.2260	.000	.	5840

6.3. Análisis clúster no jerárquico: método de k-modas

Como ya tenemos el número óptimo de grupos, el siguiente paso es llevar a cabo el análisis clúster no jerárquico con el **conjunto total de la muestra inicial** ya que este algoritmo de agrupación no genera problemas computacionales con conjuntos grandes de datos [14]. Tal y como mencionamos en la metodología, este paso lo llevaremos a cabo con el uso del software R.

La sentencia empleada es la siguiente. Basta con llamar al conjunto de datos que contenga las seis variables que estamos empleando, e indicar el número de grupos escogidos, en nuestro caso 10.

```
cluster <- kmodes(todos, 10, iter.max = 400, weighted = FALSE, fast = TRUE)
```

La salida de esta función es el clúster al que cada observación ha sido asignada. Por lo que debemos unir esta salida con el conjunto de datos empleado para la realización del clustering.

```
nueva <- cbind(todos2, CLUSTER=cluster$cluster)
```

Antes de importar el archivo generado, modificamos el nombre de una serie de categorías que se han visto afectadas por la importación del archivo desde SAS a R. En el anexo podemos encontrar las correcciones realizadas. Tras aplicar la depuración pertinente a este conjunto de datos, los exportamos para continuar trabajando en SAS Base, y proceder a su caracterización.

6.4. Caracterización de los clústeres del perfil académico

Una vez tenemos el archivo importado en SAS con las observaciones asignadas a los clústeres obtenido en el paso anterior, uniremos este archivo con el conjunto total de datos, y a continuación procederemos a caracterizarlos. La caracterización la realizaremos sobre el conjunto de variables utilizadas para la formación de los clústeres.

- **Variables binarias**

De las seis variables con las que estamos trabajando, todas excepto la rama son binarias. Recordemos que para este apartado crearemos un mapa de calor. En el eje horizontal mostraremos una categoría de referencia, y en el vertical cada uno de los clústeres formados.

Haremos uso de una leyenda de color donde el color verde significará una baja presencia de individuos de esa categoría en ese clúster, y, por el contrario, el azul una alta representación. El color blanco indicará que no existe una predominancia de ninguna de las categorías. Los aspectos más llamativos son los cruces que se encuentran en los dos extremos, es decir los colores verdes y azules oscuros.

El procedimiento para la preparación del archivo que utilizaremos para graficar, lo hemos realizado en SAS Base, y consiste en los siguientes pasos:

1. A través de un *proc freq* exportamos una tabla de frecuencias con las variables seleccionadas por clúster.
2. Mediante el procedimiento *proc transpose* transponemos la matriz resultante del paso anterior y seleccionamos una de las dos categorías de la variable, eliminando, por tanto, la restante.

Las variables binarias utilizadas para esta caracterización son las recogidas en la Tabla 8, donde la segunda columna indica la categoría de referencia escogida para cada variable.

Una primera visión que obtenemos del Gráfico 16 son el clúster 2 y 7 que están formados al 100% por mujeres. El clúster 10 está formado por estudiantes de universidad pública en su totalidad, y son el 6 y el 9 los que corresponden a estudiantes, en su mayoría, de universidad privada. En cuanto a si estos han recibido beca, los clústeres 4 y 7 sí lo han hecho, mientras que los individuos del clúster 10 no lo han hecho.

Tabla 8. Variables utilizadas para la formación del mapa de calor de los perfiles académicos

VARIABLE	VARIABLE	CATEGORÍA SELECCIONADA
SEXO	Sexo	Mujer
T_UNIV	Tipo de Universidad	Universidad Pública
EST_B1	Disfrutó de alguna beca durante sus estudios	Sí
EST2_MA	Ha completado un máster en una universidad española	Sí
IDIOMA_OFICIAL	Posee título oficial de idiomas	Sí

Las dos siguientes variables no muestran colores tan oscuros, pero aun así se aprecian aspectos relevantes. Los individuos de los clústeres 3, 6, 8 y 10 en su mayoría han estudiado un máster y coincide con los que tienen el idioma acreditado, además del 9.

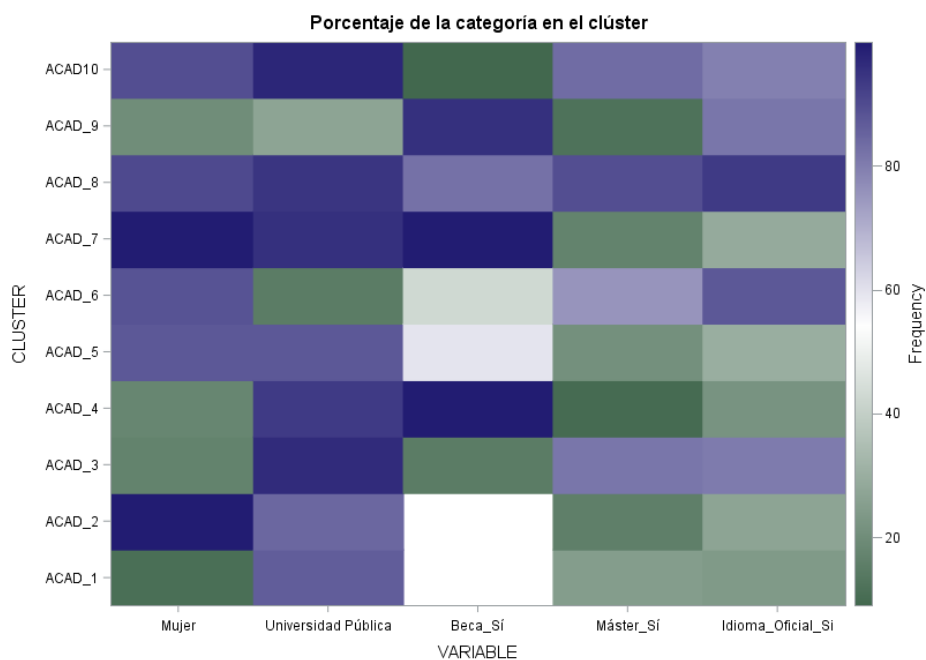


Gráfico 16. Mapa de calor para la caracterización de los perfiles académicos.

- **Variables de más de dos categorías**

En este caso, solo se trata de la variable que recoge la rama de la titulación estudiada. Por lo tanto, enfrentaremos dicha variable con la que contiene los clústeres en un análisis de correspondencias simples, ya detallada en el apartado de metodología.

El código SAS utilizado para comenzar con el ACS es el siguiente.

```
proc corresp data=trabajo.clusterkmodas all chi2p print=both DIMENS=3;
tables RAMA, Cluster;
run;
```

El objetivo del ACS realizado es comprobar si existe relación entre los clústeres formados y la variable *rama*. En la figura que mostramos a continuación, vemos la descomposición de la inercia y del estadístico Chi-Cuadrado utilizado para el contraste de independencia entre variables. Comprobamos que el *p-valor* es menor a 0.01, por lo que rechazamos

la hipótesis nula de independencia de variables. Así que, podemos continuar con el análisis al comprobar que existe relación de dependencia.

También del Gráfico 17, deducimos que retendremos tres dimensiones, ya que con 2, tras comprobar la calidad de la representación vimos que era inferior al 50% en alguna de las categorías, y, además, con 3 conseguimos explicar prácticamente el 97% de la variabilidad de los datos.

Puesto que con tres dimensiones obtenemos tres gráficos, y por cuestiones de espacio en la memoria y facilidad de interpretación, haremos uso de las tablas ofrecidas por el procedimiento *proc corresp*. Crearemos la Tabla 9 que será un resumen a partir de las contribuciones de cada categoría a la inercia, y de los signos de las coordenadas de las mismas. Con ello, conseguiremos encontrar las categorías (tanto de filas como de columnas) que mayor aportación hacen a la inercia de cada dimensión

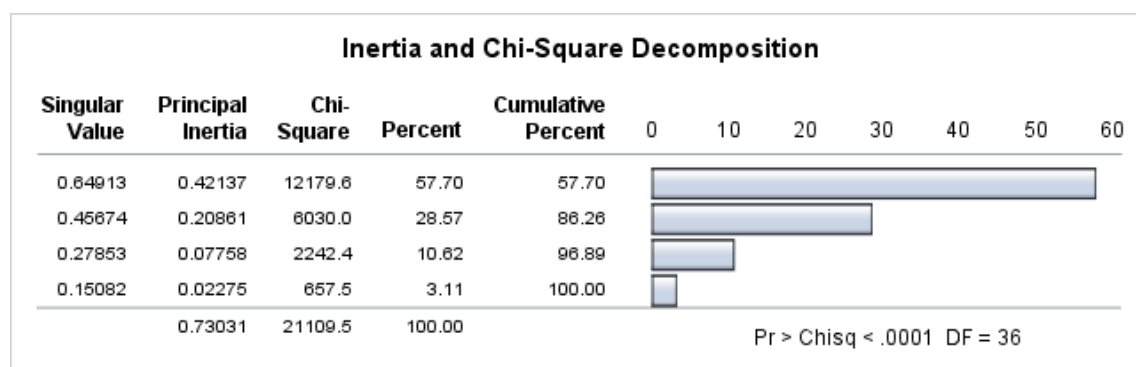


Gráfico 17. Descomposición de la inercia del ACS de la variable rama de estudio de los perfiles académicos

En dicha Tabla 9, la primera columna indica la dimensión en cuestión, en la segunda la contribución de la categoría a la inercia, en la tercera la categoría en cuestión y el signo de la coordenada sobre los ejes. Lo mismo en las columnas siguientes para los clústeres. La mínima contribución a la inercia la establecemos en 0.6. Esto nos permite identificar las ramas y clústeres que permitirán explicar mejor cada dimensión. En la última columna encontramos la relación que existe entre los signos de las categorías de las filas y columnas. Si los signos de las coordenadas coinciden, esta será directa, y si son opuestos, inversa. Nos apoyaremos del Gráfico 18 que enfrenta la dimensión uno y dos, para detectar relaciones entre las categorías diferentes.

Siguiendo la Tabla 9, la dimensión uno es explicada en gran parte por la rama de ingeniería y arquitectura, y por los clústeres 1 y 4. La relación entre ambas categorías es directa, por lo que podríamos decir que la rama de ingeniería y arquitectura está asociada con los clústeres 1 y 4.

La dimensión dos está explicada en su mayoría por la rama de artes humanidades y el clúster 8. La relación entre ambos es directa. En cuanto a la dimensión tres, se explica en su mayoría por la rama de ciencias y el clúster 10. De nuevo, esta relación es directa.

Tabla 9. Resumen de las principales aportaciones a la inercia entre la rama de estudio y los clústeres del perfil académico

Dim.	Contribución parcial a la inercia	Categoría	Signo coord.	Contribución a la inercia	Clúster	Signo coord.	Relación
DIM 1	0.7628	Ingeniería y arquitectura (a)	Positivo	0.3135	1	Positivo	Directa(a,1)
				0.3264	4	Positivo	Directa (a,4)
DIM 2	0.6552	Artes y humanidades	Positivo	0.6037	8	Positivo	Directa
DIM 3	0.7324	Ciencias	Positivo	0.6707	10	Positivo	Directa

La dimensión dos está explicada en su mayoría por la rama de artes humanidades y el clúster 8. La relación entre ambos es directa. En cuanto a la dimensión tres, se explica en su mayoría por la rama de ciencias y el clúster 10. De nuevo, esta relación es directa.

Si observamos el Gráfico 18, donde se enfrenta la dimensión uno a la dos, vemos que con ciencias de la salud se asocian los clústeres 2 y 7, y con ciencias sociales el 5 y 9. Además, observamos que la dimensión uno separa en la parte del signo positivo a la rama de ingeniería y arquitectura, mientras que las demás se encuentran en la parte de signo negativo.

En cuanto a la dimensión dos, podríamos decir que ordena, desde la parte positiva a la negativa, el nivel (de mayor a menor) de *abstracción* de las ramas representadas sobre los ejes.

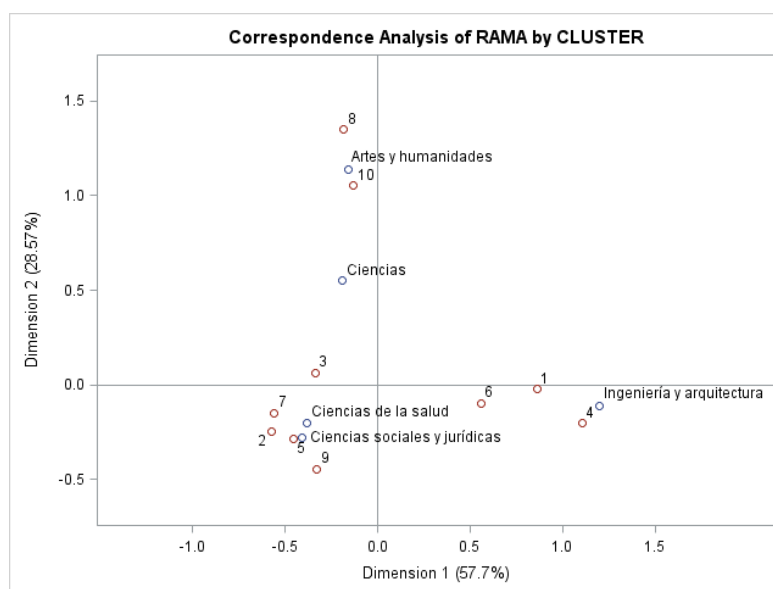


Gráfico 18. Representación de las modalidades sobre los ejes 1 y 2 del ACS de la rama de estudio y el perfil académico

Dado que la rama de ciencias es explicada por la dimensión dos recogeremos las representaciones de esta dimensión frente a la dimensión uno y dos, los cuales pueden

consultarse en el Anexo. De ellos confirmamos que dicha rama se relaciona con el clúster 10 dado el ángulo pequeño con respecto al origen.

6.5. Resumen de la caracterización y descripción de los perfiles académicos

Una vez hemos caracterizado las variables por separado, recogemos en Tabla 10 un resumen de las características principales de cada clúster con respecto a las variables.

Tabla 10. Resumen de la caracterización del perfil académico

Clúster académico	Sexo	Tipo de universidad	Rama	Máster	Beca	Idioma acreditado	Frecuencia
CLUSTER 1	Hombre	Pública	Ingeniería y arquitectura	No	-	No	17.75
CLUSTER 2	Mujeres	Pública	Ciencias de la salud	No	-	No	16.66
CLUSTER 3	Hombres	Pública	Ciencias sociales y jurídicas*	Sí	No	Sí	10.73
CLUSTER 4	Hombres	Pública	Ingeniería y arquitectura	No	Sí	No	11.23
CLUSTER 5	Mujeres	Pública	Ciencias sociales y jurídicas	No	-	No	9.41
CLUSTER 6	Mujeres	Privada	Ingeniería y arquitectura	Sí	-	Sí	2.97
CLUSTER 7	Mujeres	Pública	Ciencias de la salud	No	Sí	No	15.12
CLUSTER 8	Mujeres	Pública	Artes y humanidades	Sí	Sí	Sí	6.89
CLUSTER 9	Hombres	Privada	Ciencias sociales y jurídicas	No	Sí	Sí	5.07
CLUSTER 10	Mujeres	Pública	Ciencias	Si	No	Sí	4.15

*rama mayoritaria

Para enriquecer esta caracterización buscaremos cuales son las titulaciones más frecuentes en cada clúster según la rama asignada. Para ello recurriremos a R donde programaremos un bucle. Escogeremos las 3 titulaciones más frecuentes para cada caso, que recogemos en la Tabla 11.

A continuación, mostramos los perfiles obtenidos gracias al análisis anterior haciendo énfasis en sus características más diferenciadoras.

Perfil 1: Formado por dos clústeres de mujeres estudiantes de universidad pública de la rama de ciencias de la salud que difieren entre sí en haber recibido beca o no.

Perfil 2: Mujeres estudiantes de universidad pública de titulaciones relacionadas con las ciencias sociales.

Perfil 3: Mujeres estudiantes de universidad pública de titulaciones relacionadas con estudios ingleses e historia de la rama de artes y humanidades.

Perfil 4: Mujeres estudiantes de universidad pública de titulaciones relacionadas con las ciencias naturales de la rama de ciencias.

Perfil 5: Mujeres estudiantes de universidad privada de titulaciones relacionadas con arquitectura y edificación.

Perfil 6: Formado por dos clústeres de hombres estudiantes de universidad pública de la rama de ingeniería y arquitectura que difieren entre sí en haber recibido beca o no.

Perfil 7: Hombres estudiantes de universidad privada de titulaciones relacionadas con el derecho y la administración y gestión de empresas.

Tabla 11. Principales titulaciones de los clústeres de los perfiles académicos

CLÚSTER ACADÉMICO	RAMA	TITULACIONES
CLUSTER 1	Ingeniería y arquitectura	Ing. Técnica Industrial Grado en Ing. De la Edificación Ing. Téc. Agrícola
CLUSTER 2	Ciencias de la salud	Dip. En Enfermería Lic. En Medicina Dip. En Fisioterapia
CLUSTER 3	Ciencias sociales y jurídicas*	Lic. En Ciencias de la actividad Física y el deporte Lic. En Derecho Lic. En Economía
CLUSTER 4	Ingeniería y arquitectura	Ing. Técnica Industrial Ing. En Informática Ing. Téc. Agrícola
CLUSTER 5	Ciencias sociales y jurídicas	Dip. En Trabajo Social Maestro-Especialidad Educación Infantil Lic. En Psicología
CLUSTER 6	Ingeniería y arquitectura	Grado en Ing. De la Edificación Arquitecto Arquitecto Técnico
CLUSTER 7	Ciencias de la salud	Dip. En Enfermería Lic. En Medicina Dip. En Fisioterapia
CLUSTER 8	Artes y humanidades	Lic. En Filología Inglesa Lic. En Traducción e Interpretación Lic. En Historia
CLUSTER 9	Ciencias sociales y jurídicas	Lic. En Derecho Lic. En Administración de Empresas Dip. En ciencias empresariales
CLUSTER 10	Ciencias	Lic. en Biología Lic. En Química Lic. En Ciencias Ambientales

*rama mayoritaria

7. Clasificación de los grupos de perfiles laborales

El siguiente objetivo que perseguimos es determinar los perfiles laborales de la población muestreada. Por lo tanto, vamos a proceder de la misma forma que para el perfil académico.

Por motivos de la estructura de la encuesta, las preguntas realizadas a los individuos en relación con su vida laboral, van orientadas según sea su situación actual. Esta puede ser: trabajando, desempleado e inactivos. Por ello, no podemos utilizar las mismas variables para estos tres grupos. Así que hemos decidido proceder de la siguiente manera (Figura 1), haciendo una unión final de clústeres:

1. Tratar cada grupo de forma separada en el análisis clúster.
2. Escoger variables específicas para cada uno de los grupos:
 - Trabajando: seleccionaremos variables específicas para ellos y hallaremos los grupos que se generan tras el análisis clúster dentro de éstos.
 - Desempleado: dentro de estos, haremos una distinción entre aquellos que no han trabajado nunca, y los que, por el contrario, si lo han hecho. Para el caso de los que no han trabajado nunca, les consideraremos como único grupo dentro de los desempleados, sin realizar un análisis clúster sobre ellos.

En el caso de los que sí han trabajado, ya que son un porcentaje mayor, realizaremos un análisis clúster escogiendo variables que recogen información exclusivamente para ellos.

- Inactivos: Puesto que se trata del 6,7% del total de los datos, a estos vamos a considerarles como un único grupo.

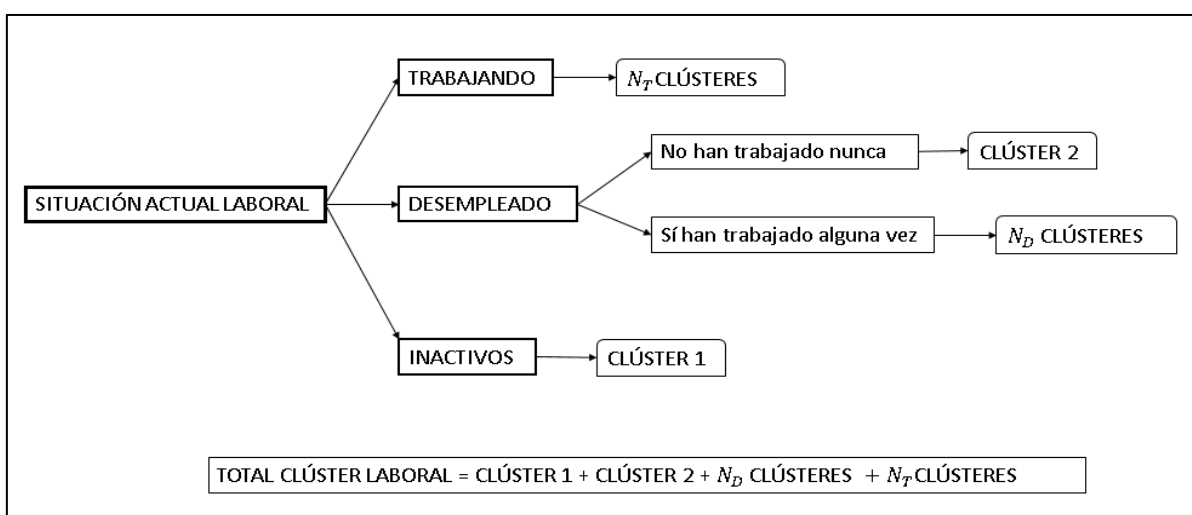


Figura 1. Descomposición para la creación de los clústeres de perfiles laborales

El procedimiento para los que se encuentran trabajando y los desempleados que sí han trabajado alguna vez, consistirá en los mismos pasos que para los clústeres del ámbito académico.

Crearemos subconjuntos de la muestra según su situación laboral, y emplearemos el correspondiente a cada punto.

7.1. Población que se encuentra trabajando

El conjunto formado por personas trabajando está formado 21.484 observaciones.

7.1.1. Selección de variables

- **Selección de una muestra aleatoria simple de tipo instrumental**

De nuevo, el número de observaciones con las que cuenta este subconjunto de la muestra no nos permite realizar el análisis clúster jerárquico dado la longitud de los datos. Por lo tanto, debemos recurrir a un muestreo aleatorio simple de 20.000 observaciones de entre las 21.484 que componen este grupo de personas trabajando. El motivo por el cual no realizamos un muestreo estratificado es que la variable de estrato es la que utilizamos ahora para obtener los subconjuntos laborales.

Una vez tenemos conjunto de datos con el que podemos realizar la selección de variables procedemos a ello. Las variables disponibles con las que realizaremos las pruebas para obtener el conjunto de ellas que mejores resultados ofrece son:

- País en el que desempeña su labor profesional
- Situación profesional actual
- Tipo de jornada de trabajo actual
- Ocupación que desempeña en el puesto de trabajo actual
- Actividad principal del establecimiento o local donde trabaja actual
- Número de personas trabajando en el establecimiento o local donde trabaja actualmente
- Año en el que empezó a trabajar en este último empleo
- Base de cotización a la SS en marzo de 2014 de los empleados por cuenta ajena (Quintiles)

Depuración del subconjunto

Todas las agrupaciones que realicemos en subconjuntos las trasladaremos al conjunto total de datos, pero la necesidad de realizarlo la descubrimos según la necesidad de uso de la variable, ya que contamos con 180 en el conjunto inicial.

Comenzamos con la variable *TR_ANIO* la hemos agrupado en cinco categorías ya que presentaba información desde 1965, y muchos de ellos solo tenían a un individuo. Por lo que la agrupación se queda como vemos en la Tabla 12.

Tabla 12. Reagrupación de la variable "año en el que empezó a trabajar en su actual empleo"

Año en el que empezó a trabajar en este último empleo				
TR_ANIO	Frequency	Percent	Cumulative Frequency	Cumulative Percent
2010 o menos	4219	19.64	4219	19.64
2011	2797	13.02	7016	32.66
2012	2999	13.96	10015	46.62
2013	5145	23.95	15160	70.56
2014 o más	6324	29.44	21484	100.00

Otra variable que hemos agrupado es *LTR_PAÍ* que indica el país en el que está trabajando el individuo. Dada la desigual frecuencia entre categorías, hemos decidido agruparla en tres categorías, como muestra la Tabla 13.

Tabla 13. Reagrupación de la variable "país en el que está trabajando" el individuo

LTR_PAÍ				
LTR_PAÍ	Frequency	Percent	Cumulative Frequency	Cumulative Percent
España	20251	70.06	20251	94,26
Fuera de Europa	551	1.91	20802	96,83
Otro país europeo	682	2.36	21484	100

También, al igual que hicimos en el clúster laboral en el caso de la titulación, eliminaremos de nuestras posibilidades la variable TRABOC, ya que presenta más de 40 categorías.

En cuanto a la variable que recoge la información de la base de cotización a la SS, hemos escogido la de 2014, ya que tratamos de caracterizar el trabajo que tenían los individuos en el momento de la encuesta, año 2014.

Una vez que tenemos estas variables agrupadas, podemos continuar con la selección de variables, de la misma forma que en el primer apartado. La Tabla 14 muestra un resumen de las pruebas realizadas. Para información más detallada, puede consultarse el anexo, donde recogemos los resultados de los mejores resultados para cada prueba.

Tabla 14. Resultado de las pruebas de selección de variables para perfiles laborales de individuos trabajando

Prueba	Mejor opción de cada prueba	
	Número de clústeres	Variabilidad explicada
Pruebas con 9 variables	30	55.7%
Pruebas con 8 variables	30	62.3%
Pruebas con 7 variables	30	63.2%
Pruebas con 6 variables	23	70.4%
Pruebas con 5 variables	13	70.2%

Finalmente, tras realizar las pruebas pertinentes, las variables que mejores resultados ofrecen son las de la Tabla 15.

Tabla 15. Variables seleccionadas para obtener los perfiles laborales de individuos trabajando

Variable	Categorías
SEXO	Hombre/Mujer
SIT_PRO	Asalariado con contrato permanente/Asalariado con contrato temporal/Ayuda familiar/Empresario o trabajador independiente/En prácticas, formación o becario/No aplicable
JORNADA	A tiempo completo / A tiempo parcial
TR_CNAE	* 32 niveles. Actividad principal del establecimiento o local donde trabaja actual
BCOTC14	Primer quintil/Segundo quintil/Tercer quintil/Cuarto quintil/Quinto quintil/No aplicable

*Anexo

7.1.2. Análisis clúster jerárquico

Una vez hemos determinado el conjunto de variables podemos proceder a determinar el número de grupos óptimo.

- **Matriz de distancias**

Para el cálculo de la matriz de distancias emplearemos la siguiente sentencia SAS. En ella debemos indicar que el método escogido es *djaccard* y las variables seleccionadas.

```
proc distance data=trabajo.trabajandostra method=djaccard absent=0  
out=trabajo.distjaclab;  
var anominal(SEXO SIT_PRO JORNADA TR_CNAE BCOTC14 );  
id IDEN;  
run;
```

- **Determinar el número de grupos**

Una vez tenemos el archivo que utilizaremos de entrada para el clúster jerárquico, realizaremos tres pruebas para obtener el mejor método de cálculo de distancias, al igual que para el clúster académico.

Los resultados de dichas pruebas se recogen en la Tabla 16, donde vemos que el método que más variabilidad consigue explicar con un menor número de agrupaciones es el de Ward.

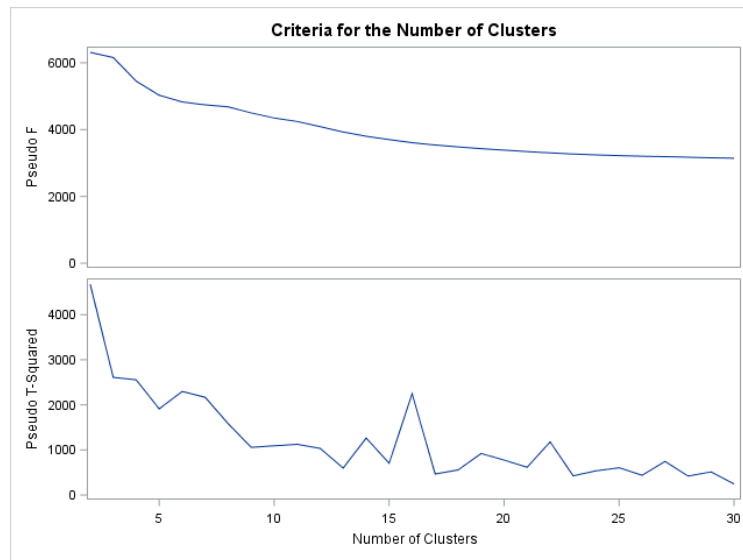
Tabla 16. Resultados de la elección del método de cálculo de distancias entre clústeres de los perfiles laborales de individuos trabajando

Métodos de cálculo de distancias	Agrupaciones	Variabilidad explicada
Enlace medio	14	65,8%
Distancia entre centroides	12	61,5%
Distancia de Ward	9	64%

Por lo tanto, procedemos a ejecutar el procedimiento *proc cluster* con el método de Ward, y analizar con más detalle los resultados ofrecidos.

Dado que el clúster laboral se conforma por la suma de grupos de diferentes orígenes según su situación laboral, debemos ser prudentes en cuanto a la elección del número de clúster para los que están trabajando y los desempleados que si han trabajado alguna vez. Por ese motivo, buscaremos que los grupos expliquen al menos el 60% de la variabilidad total. A partir de ese corte, analizaremos los estadísticos *Pseudo-t squared* y *Pseudo-F statistics*.

A continuación, vemos el Gráfico 19 que muestra la evolución de estos estadísticos a medida que se incrementa el número de clústeres. En él observamos como en el clúster número 8 existe un máximo relativo del *Pseudo-t Squared*, y a continuación un descenso pronunciado del mismo. Si recordamos lo explicado en el clúster académico, a este número debemos sumarle uno. Además, con 9 grupos conseguiríamos explicar el 64,3% de la variabilidad de los datos. Así que procederemos el análisis con 9 grupos. La tabla de la que procede esta información puede consultarse en el anexo.



En el Gráfico 19 observamos como en el clúster número 8 existe un máximo relativo del *Pseudo-t Squared*, y a continuación un descenso pronunciado del mismo. Si recordamos lo explicado en el clúster académico, a este número debemos sumarle uno. Además, con 9 grupos conseguiríamos explicar el 64,3% de la variabilidad de los datos. Así que procederemos el análisis con 9 grupos. La tabla de la que procede esta información puede consultarse en el anexo.

7.1.3. Análisis clúster no jerárquico: método de k-modas

Una vez hemos determinado el número de grupos, procederemos a realizar el análisis clúster no jerárquico con el algoritmo de agrupación *k-modas* con el conjunto total de los 21.484 individuos que conforman esta subpoblación.

Para ello, leemos el archivo en R, y aplicamos la función correspondiente de la librería, donde como ya sabemos, debemos señalar el número de grupos escogidos y sobre las variables que queremos ejecutar el algoritmo.

Cuando este proceso ha terminado, depuramos los datos puesto que en este proceso el nombre de las categorías sufre alguna transformación. Unimos el archivo generado por esta función, de donde solo seleccionaremos la columna perteneciente al clúster, al archivo que contiene todas las variables. Finalmente lo exportamos a Excel y posteriormente lo importamos en SAS Base.

7.2. Población desempleada que sí ha trabajado alguna vez

El siguiente grupo con el que vamos a obtener el número de clúster óptimo es con los individuos que se encuentran en desempleo, pero que sí han trabajado alguna vez. Procederemos de la misma forma que para los casos anteriores.

Las variables disponibles para este subconjunto de la población y con las que realizaremos las pruebas de selección de variables se encuentran en la Tabla 17.

Tabla 17. Variables disponibles para los individuos en desempleo

Variables	Descripción	Categorías
SEXO	Género	Hombre Mujer
BQ_NUM	Número de empleadores distintos que ha tenido	1/2/3/4 o más
BQ_TP_P	Tiempo que ha estado trabajando desde que terminó los estudios	Menos de 6 meses/De 6 meses a 1 año/De 1 año a año y medio/De 1 año y medio a 2 años/2 o más años
BSQRE	Ha rechazado algún trabajo por considerarlo poco adecuado desde que se ha titulado	Sí No
DISTRE6	Utilidad de su titulación para encontrar trabajo.	Sí No

7.2.1. Selección de variables

Tras realizar las pruebas, hemos decidido escoger 4 variables ya que este número ofrece los mejores resultados. En la Tabla 18 vemos las pruebas que mejores resultados ofrecieron.

Tabla 18. Resultado de las pruebas de selección de variables para perfiles laborales de individuos desempleados

Prueba	Número de grupos	Variabilidad explicada
Prueba 1 con 5 variables	11	71.1%
Prueba 2 con 4 variables	9	71.2%

Las cuatro variables seleccionadas en las pruebas anteriores se encuentran en Tabla 19:

Tabla 19. Variables seleccionadas para obtener los perfiles laborales de individuos desempleados

VARIABLE	DESCRIPCIÓN
SEXO	Género
BQ_NUM	Número de empleadores distintos que ha tenido
BQ_TP_P	Tiempo que ha estado trabajando desde que terminó los estudios
TRAB_F1	Ha tenido algún trabajo remunerado durante la realización de los estudios

7.2.2. Análisis clúster jerárquico

- **Matriz de distancias**

En este caso no es necesario el paso de seleccionar una muestra estratificada, ya que estamos tratando con 5 mil observaciones aproximadamente, y no nos genera problemas de coste computacional. Así que procederemos directamente a hallar la matriz de distancias con el coeficiente de Jaccard.

- **Determinar el número de grupos**

De nuevo realizamos las pruebas pertinentes para hallar el mejor método de cálculo de distancias. Los resultados de las pruebas con más detalle pueden consultarse en el anexo, y a continuación mostramos un resumen de las mismas en la Tabla 20.

El método de enlace medio y distancia de centroides ofrece la misma variabilidad explicada para el mismo número de agrupaciones. A pesar de eso, el método escogido es de enlace medio ya que, a la hora de determinar el número de grupos, el gráfico del *Pseudo-t* era más claro y determinante. Mientras que el del otro método podría llevar a confusión.

Tabla 20. Resultados de la elección del método de cálculo de distancias entre clústeres de los perfiles laborales de individuos desempleados

Métodos de cálculo de distancias	Agrupaciones	Variabilidad explicada
Enlace medio	4	60,2%
Distancia entre centroides	4	60,2%
Distancia de Ward	4	55,7%

A continuación, mostramos el Gráfico 20 asociado a los estadísticos del *Pseudo-t squared* y *Pseudo F Statistics*, y la tabla *Cluster History* podemos encontrarla en el anexo. A partir de ellos, vemos que en el clúster 3 se da un máximo relativo con un fuerte descenso en el *Pseudo-t squared*. Esto nos estaría indicando que escogiéramos 4 grupos. Además, con este número conseguimos explicar el 60,2% de los datos.

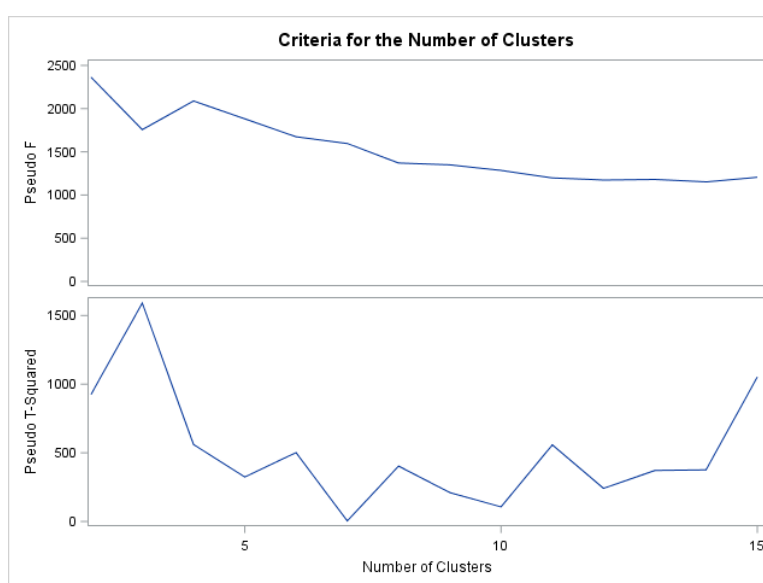


Gráfico 20. *Pseudo F* y *Pseudo T* para la selección del número de clústeres de los perfiles laborales de individuos desempleados

7.2.3. Análisis clúster no jerárquico: método de *k-modas*

El procedimiento llevado a cabo para realizar en análisis clúster no jerárquico es el mismo que se viene haciendo hasta ahora. Llevamos el archivo de desempleados a R, donde aplicaremos la función del algoritmo *k-modas*. Realizamos la depuración pertinente y lo importamos en SAS.

7.3. Obtención de los clústeres totales del ámbito laboral

Finalmente, la agregación de todos los clústeres del ámbito laboral que hemos ido obteniendo de forma desagregada en los pasos anteriores, conforman un total de 15 clústeres. En la Tabla 21 vemos de forma resumida la contribución de cada población de estudio, y su aportación al grupo final.

Así que, una vez tenemos todos los clústeres de la parte laboral, nos queda unir los archivos generados en cada paso. Obtendremos el conjunto total de datos, donde cada individuo pertenecerá a uno de los 15 clústeres laborales.

Tabla 21. Agregación de los clústeres obtenidos de todos los perfiles laborales

CLÚSTER	Número de clústeres	Clústeres asignados
Clúster inactivo	1	1
Clúster desempleado. No han trabajado nunca	1	2
Clúster desempleado. Sí han trabajado	4	3,4,5,6
Clúster trabajando	9	7,8,9,10, 11, 12, 13, 14, 15
CLUSTER LABORAL	15	

7.4. Caracterización de los clústeres conformados del ámbito laboral

La caracterización del clúster laboral la haremos por partes, de la misma forma que para su creación. En primer lugar, procederemos a describir al clúster conformado por los que se encuentran actualmente trabajando. El procedimiento será igual al realizado para los clústeres del perfil académico.

7.4.1. Caracterización de la población que está trabajando

Las variables que emplearemos para la caracterización son las mismas que utilizamos en el análisis clúster. A excepción de los inactivos y los desempleados que no han trabajado nunca, que los que emplearemos variables específicas consideradas en la encuesta.

- **Variables binarias**

Son dos variables binarias con las que contamos en este caso, el sexo del individuo y el tipo de jornada de trabajo. El procedimiento para obtener los datos necesarios y de la forma adecuada para crear el mapa de calor (Gráfico 21) será igual que en el caso anterior.

Destacamos una mayor presencia de mujeres en los clústeres 7, 9, 13 y 14, siendo el clúster 7 el que tiene un mayor %. Así como de hombres en los 8, 10, 11 y 15. El clúster 12 está formado igualmente por hombres como por mujeres.

En cuanto al tipo de jornada laboral, únicamente el clúster 9 está asociado con una jornada a tiempo parcial. A excepción de los clústeres 9, en el resto predomina la jornada a tiempo completo. En el 12 y 13 no queda claro un perfil.

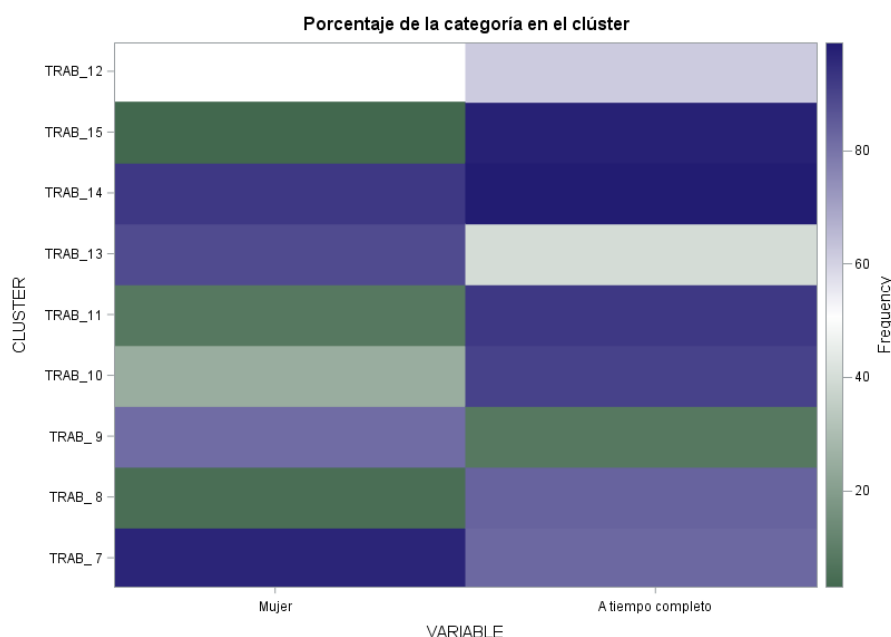


Gráfico 21. Mapa de calor para la caracterización de los perfiles laborales de individuos trabajando

- **Variables con más de dos categorías**

Las variables con las que realizaremos el análisis se encuentran en Tabla 22. Analizaremos en cada uno la hipótesis de independencia frente a la variable que clúster, así como las dimensiones necesarias para alcanzar aproximadamente al menos el 90% de la variabilidad total.

Tabla 22. Conjunto de variables con más de dos categorías de los perfiles laborales de individuos trabajando

Variable	Descripción	Número de categorías
SIT_PRO	Situación profesional actual	6
TR_CNAE	Actividad principal del establecimiento o local donde trabaja actual	32
BCOTC14	Base de cotización a la SS en marzo de 2014 de los empleados por cuenta ajena (Quintiles)	6

- **Base de cotización a la SS en 2014 (BCOTC14) frente a los clústeres**

La primera variable que caracterizaremos es la base de cotización a la SS en marzo de 2014 de los empleados por cuenta ajena, expresada en quintiles.

En el Gráfico 22 aparece el p-valor del contraste Chi-Cuadrado que nos lleva a rechazar la hipótesis de independencia y así, continuar con el análisis. Además, si nos fijamos en la última columna vemos como con cuatro dimensiones logramos alcanzar el 93% de la variabilidad total.

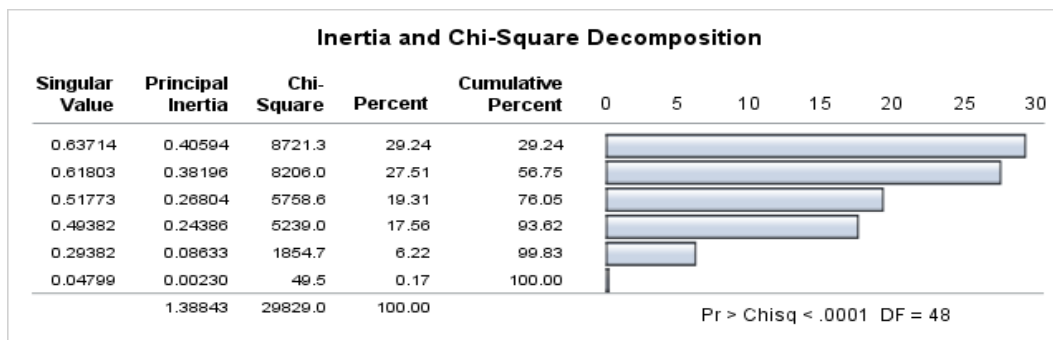


Gráfico 22. Descomposición de la inercia del ACS de la base de cotización a la Seguridad Social de los perfiles laborales de individuos trabajando

Tabla 23. Calidad, masa e inercia de los clústeres del perfil laboral de individuos trabajando del ACS con la base de cotización a la Seguridad Social en 2014

Summary Statistics for the Column Points			
	Quality	Mass	Inertia
7	0.9688	0.3297	0.0497
8	0.9865	0.1639	0.0472
9	0.4875	0.1776	0.0788
10	0.9542	0.0966	0.2101
11	0.9990	0.0864	0.2208
12	0.9714	0.0288	0.0240
13	0.9836	0.0417	0.1867
14	0.9519	0.0529	0.1523
15	0.9887	0.0223	0.0303

Previamente a la obtención e interpretación de los resultados, a través de la Tabla 23 confirmaremos que con cuatro dimensiones obtenemos una buena calidad de los clústeres. A excepción del clúster 9, los demás se sitúan por encima del 95%, por lo que esto nos confirma que el número de dimensiones es adecuado.

Al igual que en el ámbito académico, utilizaremos la misma tabla resumen que mostramos a continuación (Tabla 24), y destacaremos algunos aspectos relevantes. Todas las tablas de las que procede esta información pueden consultarse en el anexo.

- El clúster 11 está relacionado directamente con el primer quintil de la base de cotización a la SS, y ambos explican la dimensión uno.
- La dimensión dos está explicada por la categoría no aplicable, correspondiente a los autónomos, y por los del quinto quintil. También por los clústeres 10 y 11. La única relación directa que entramos es entre el quinto quintil y el clúster 10 y 11.
- La dimensión tres viene explicada por el segundo quintil y el clúster 13 cuya relación entre ambos es directa.
- Por último, el cuarto quintil y el clúster 14 explican la cuarta dimensión, relacionándose entre sí de forma directa.

Tabla 24. Resumen de las principales aportaciones a la inercia de las dimensiones de la base de cotización a la Seguridad Social en 2014 y los clústeres de los perfiles laborales de individuos trabajando

Dim.	Contribución a la inercia	Categorías	Signo coord.	Contribución a la inercia	Clúster	Signo coord.	Relación
DIM 1	0.634	Primer Quintil (a)	Positivo	0.3221	10	Negativo	Inversa (a,10)
				0.5353	11	Positivo	Directa (a,11)
DIM 2	0.3143	No aplicable (b)	Negativo	0.3163	10	Positivo	Inversa (b,10) Directa (c,10)
				0.2086	11	Positivo	Inversa (b,11) Directa (c,11)
	0.2991	Quinto Quintil (c)	Positivo	0.1497	13	Negativo	Directa (b,13) Inversa (c,13)
DIM 3	0.6449	Segundo Quintil	Positivo	0.7241	13	Positivo	Directa
DIM 4	0.7112	Cuarto Quintil	Positivo	0.7681	14	Positivo	Directa

- **Actividad principal del establecimiento o local donde trabaja actual (TR_CNAE) frente a los clústeres**

En este caso también podemos rechazar la hipótesis nula de independencia de variables como vemos en el Gráfico 23. Con cuatro dimensiones conseguimos explicar aproximadamente el 95% de la información.

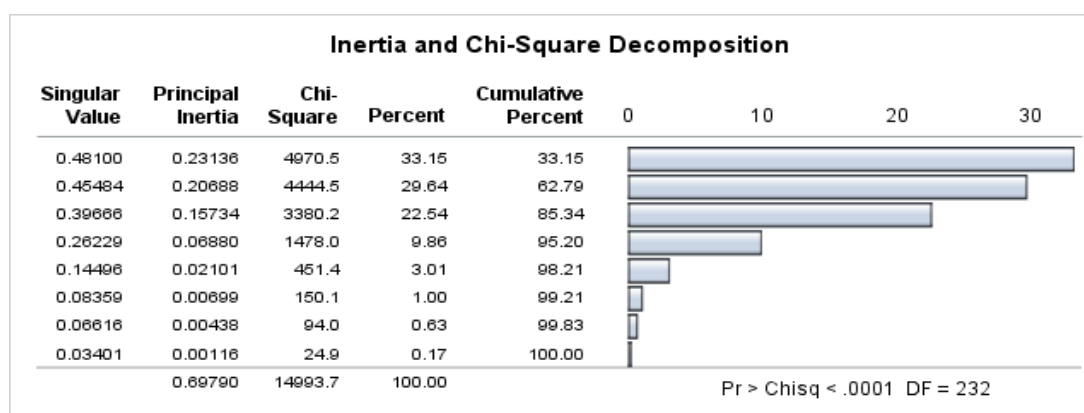


Gráfico 23. Descomposición de la inercia del ACS de la actividad principal del empleo de los perfiles laborales de individuos trabajando

La Tabla 25 nos confirma que podemos continuar con cuatro dimensiones, ya que la calidad está por encima del 80%, a excepción de los clústeres 12 y 13. De nuevo, procederemos a la caracterización mediante el uso de la información recogida en las tablas

Tabla 25. Calidad, masa e inercia de los clústeres del perfil laboral de individuos trabajando del ACS con la actividad principal del trabajo actual

Summary Statistics for the Column Points			
	Quality	Mass	Inertia
7	0.9094	0.3297	0.0968
8	0.9982	0.1639	0.2409
9	0.9904	0.1776	0.1608
10	0.9990	0.0966	0.2616
11	0.8319	0.0864	0.0679
12	0.4200	0.0288	0.0106
13	0.4666	0.0417	0.0271
14	0.9934	0.0529	0.0922
15	0.8956	0.0223	0.0420

Tabla 26. Resumen de las principales relaciones entre la actividad principal del trabajo y los clústeres del perfil laboral de individuos trabajando

Dim.	Contribución a la inercia	Categoría/s	Signo coord.	Contribución a la inercia	Clúster	Signo coord.	Relación
DIM 1	0.4989	Información y comunicaciones (a)	Positivo	0.6052	10	Positivo	Directa (a,10)
	0.3063	Educación (b)	Negativo				Inversa (b,10)
DIM 2	0.6439	Industria manufacturera	Negativo	0.7871	8	Negativo	Directa
DIM 3	0.4447	Actividades sanitarias y de servicios sociales(c)	Positivo	0.4675	9	Positivo	Directa(c,9)
	0.2303	Información y comunicaciones (d)	Negativo	0.2538	10	Negativo	Inversa (c,10)
							Directa (d,10)
DIM 4	0.6021	Administración Pública y defensa; Seguridad Social obligatoria	Positivo	0.9037	14	Positivo	Directa

A partir de la Tabla 26 destacaremos a continuación algunos aspectos:

- La categoría de información y comunicaciones se relaciona directamente con el clúster 10, tanto en la dimensión uno como en la tres.

- Las actividades sanitarias están directamente relacionadas con el clúster 9, y ambos vienen explicado en la dimensión tres.
- La dimensión cuatro muestra una relación directa con la administración pública y defensa con el clúster 14.

- Situación profesional (SIT_PRO) frente a los clústeres

En primer lugar, a partir del Gráfico 24 confirmamos que las variables son dependientes entre sí, por lo que podemos continuar con el análisis. Al conseguir explicar el 80% de los datos con dos dimensiones, nos planteamos continuar con estas.

Pero a través de las calidades de la representación sobre los ejes de los clústeres, vemos en la Tabla 27 como con dos dimensiones todos los clústeres tienen una calidad de más del 90%, mientras que el 10 tiene una calidad muy mala, en torno al 13%. En cambio, si escogemos tres dimensiones, la calidad del clúster 10 pasa al 100%. Por ello, continuaremos el análisis con tres dimensiones, consiguiendo explicar el 99,9% de la variabilidad de los datos.

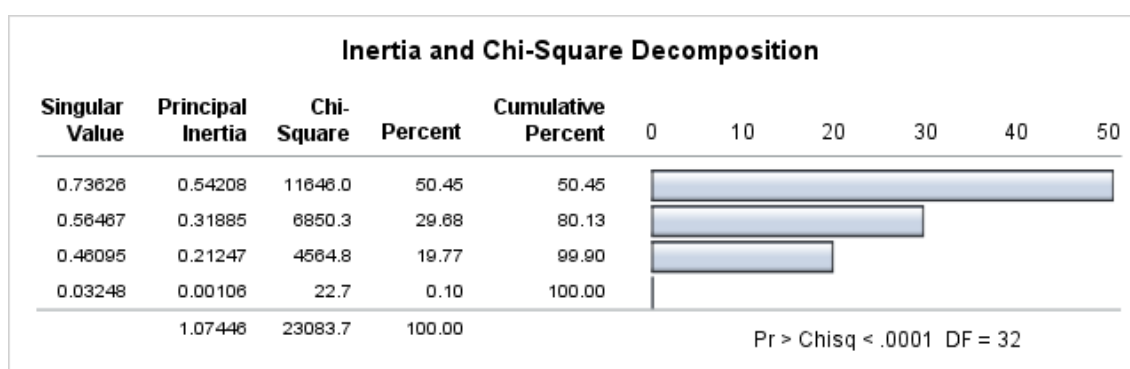


Gráfico 24. Descomposición de la inercia del ACS de la variable situación profesional de los perfiles laborales de individuos trabajando

Tabla 27. Calidades de la representación sobre 2 y 3 ejes de las modalidades de los perfiles laborales de individuos trabajando

Summary Statistics for the Column Points		
	Quality 2 dimensiones	Quality con 3 dimensiones
7	0.9818	0.9983
8	0.8782	0.9999
9	0.9771	0.9926
10	0.1372	1.0000
11	0.9753	0.9998
12	0.9331	0.9975
13	0.9995	0.9999
14	0.9893	0.9998
15	0.9963	0.9999

Puesto que hemos escogido tres dimensiones, crearemos Tabla 28 como un resumen que recoge la información de qué categorías explican cada dimensión. También nos apoyaremos del Gráfico 25.

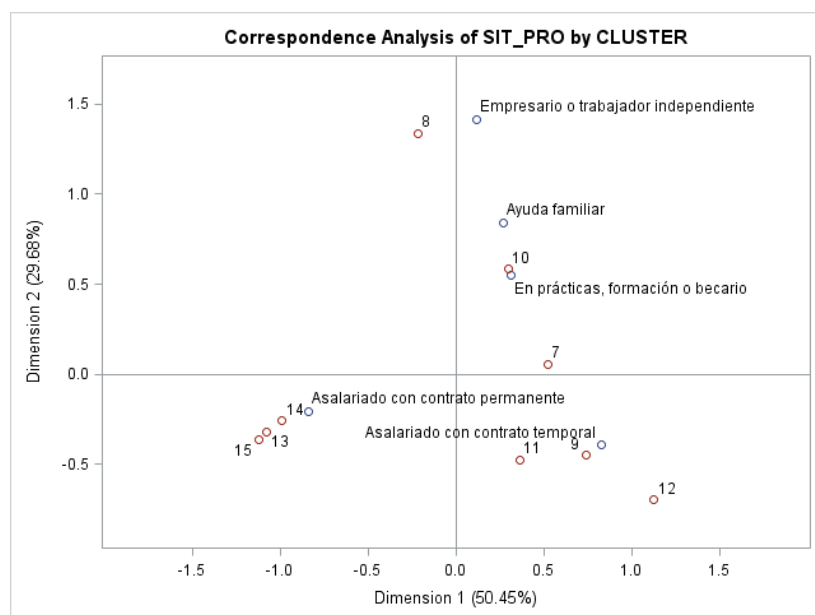


Gráfico 25. Representación de las modalidades sobre los ejes 1 y 2 del ACS de la situación profesional y los perfiles laborales de individuos trabajando

- La dimensión uno recoge información sobre la temporalidad del trabajo.
- La dimensión dos separa a los asalariados (parte del signo negativo) con los demás (parte del signo positivo).
- La dimensión uno se explica por el perfil laboral de los asalariados, la dimensión dos a los autónomos, y la 3 a los becarios o en formación.
- A través de la tabla, y lo que nos confirma el gráfico, a los asalariados con contrato permanente se les asocia directamente con los clústeres 13, 14 y 15. En cambio, los asalariados con contrato temporal están inversamente relacionados con éstos últimos clústeres, y directamente con el 9.
- A los individuos que están en prácticas, formación o becario se les relaciona con el 10. Este clúster también forma un ángulo muy pequeño con los autónomos y los de ayuda familiar. Y a los empresarios o trabajadores independientes también se les relaciona con el 8.
- Observamos que la dimensión uno contrapone a los contratos temporales en los valores positivos, y a los contratos permanentes en los valores negativos.
- Podemos apreciar en el Gráfico 25. Representación de las modalidades sobre los ejes 1 y 2 del ACS de la situación profesional y los perfiles laborales de individuos trabajando como la dimensión dos enfrenta a los autónomos (en los valores positivos) con los asalariados (en los valores negativos). Otros tipos de situaciones profesionales se encuentran en el medio de estos dos extremos.

Tabla 28. Resumen de las principales relaciones entre la situación profesional y los clústeres de los perfiles laborales de individuos trabajando

Dim.	Contribución a la inercia	Categorías	Signo coord.	Contribución a la inercia	Clúster	Signo coord.	Relación
DIM 1	0.5315	Asalariado con contrato permanente (a)	Negativo	0.1451	13	Negativo	Directa (a,13) Directa (a,14)

	0.4404	Asalariado con contrato temporal (b)	Positivo	0.2180	14	Negativo	Directa (a,15)
				0.2393	15	Negativo	Inversa (b,13) Inversa (b,14) Inversa(b,15)
DIM 2	0.6288	Empresario o trabajador independiente	Positivo	0.6481	8	Positivo	Directa
DIM 3	0.7139	En prácticas, formación o becario	Positivo	0.8141	10	Positivo	Directa

7.4.2. Caracterización de la población desempleada que sí ha trabajado alguna vez

Recordemos que para este grupo de individuos se formaron 4 clústeres. El procedimiento que llevaremos a cabo para la caracterización de este subconjunto es el mismo que en pasos anteriores.

- **Variables binarias**

En la Tabla 29 recogemos el conjunto de variables mediante las que crearemos el mapa de calor.

Tabla 29. Variables utilizadas para la formación del mapa de calor de los perfiles laborales de individuos desempleados

VARIABLE	DESCRIPCIÓN	Categoría referencia
TRAB_F1	Ha tenido algún trabajo remunerado durante la realización de los estudios	Sí
SEXO	Sexo	Mujer

A continuación, vemos el mapa de calor (Gráfico 26), donde representamos el cruce de categorías entre una categoría de cada variable y los 4 clústeres que conforman este grupo de desempleados.

Los clústeres 3 y 6 están formado por mujeres, mientras que el 4 y 5 por hombres. A excepción del clúster 3, todos los demás clústeres están formados por individuos que sí han trabajado durante sus estudios.

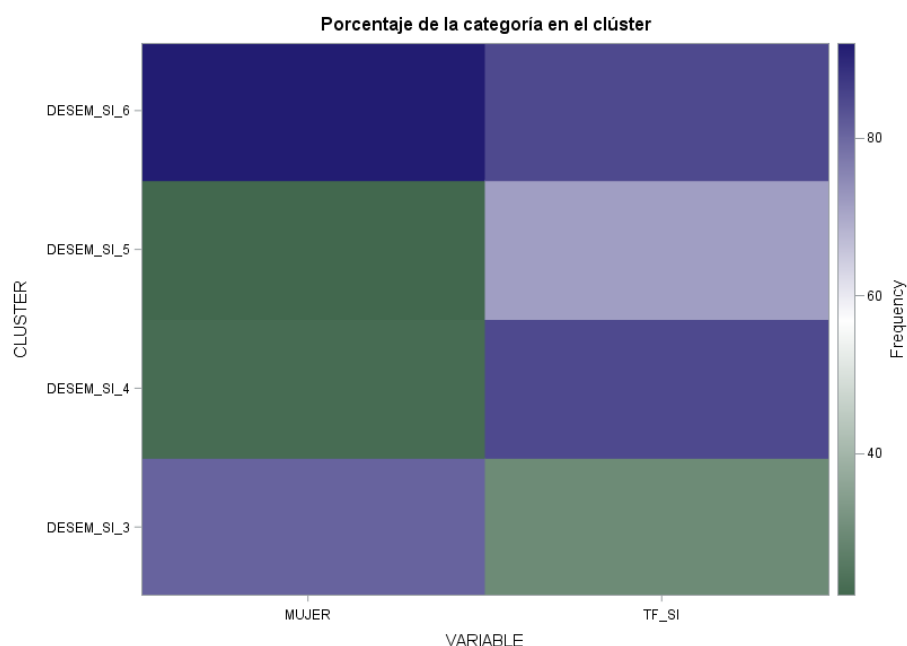


Gráfico 26. Mapa de calor para la caracterización de los perfiles laborales de individuos desempleados

- **Variables con más de dos categorías**

Las variables de más de dos categorías son las que mostramos en la Tabla 30. Procediendo con la misma mecánica, comenzaremos realizando un análisis de correspondencias simples para estas variables.

Tabla 30. Variables con más de dos categorías utilizadas para la caracterización de desempleados

Variable	Descripción
BQ_NUM	Número de empleadores distintos que ha tenido
BQ_TP_T	Tiempo que ha estado trabajando desde que terminó los estudios

- **Número de empleadores distintos que ha tenido (BQ_NUM) frente a los clústeres**

En la Gráfico 27 encontramos la descomposición de la inercia, además de rechazar la hipótesis de independencia entre variables, vemos que con tres dimensiones alcanzamos a explicar el 100%.

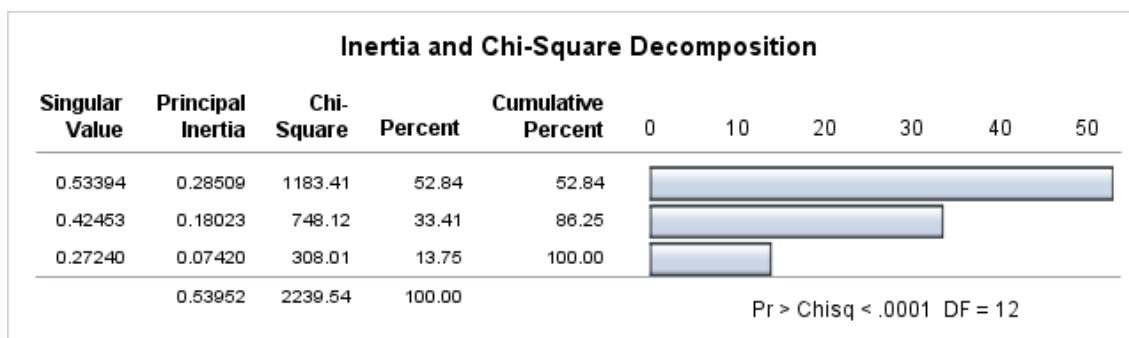


Gráfico 27. Descomposición de la inercia del ACS de la variable número de empleadores de los perfiles laborales de desempleados

Hemos escogido tres dimensiones y no dos, porque la calidad del clúster 4 es del 29%, y con 3 es del 100%, tal y como vemos en la Tabla 31

Tabla 31. Calidades de representación sobre los ejes de los clústeres al retener dos dimensiones

Summary Statistics for the Column Points			
	Quality	Mass	Inertia
3	0.7389	0.5408	0.1183
4	0.2925	0.2233	0.1335
5	0.9907	0.0990	0.3767
6	0.9765	0.1368	0.3716

Para facilitar la interpretación, hemos decidido cambiar el nombre de las categorías de *bq_num*, tal y como recogemos en la Tabla 32.

Tabla 32. Reagrupación de la variable "número de empleadores"

Nombre previo	Nombre nuevo
1	Uno
2	Dos
3	Tres
4 o más	4 o más
NS/NC	NS/NC

Realizaremos la caracterización mediante las mismas tablas utilizados en pasos anteriores, en este caso es la Tabla 33. De nuevo, establecemos como punto de corte un 60% de contribución a la inercia. También, nos apoyaremos del Gráfico 29 que contiene a la dimensión uno y dos, ya que con ellas se conseguía explicar el 86.25%. Para una información más detallada, consultar el anexo.

Las asociaciones que encontramos a través de la dimensión uno son dos, el clúster 5 está relacionado directamente con la categoría *uno*, y el clúster 6 con la categoría *tres*. Esto también nos lo confirma en Gráfico 29. Según la dimensión dos, la relación anterior también se produce al revés. Es decir, el clúster 5 con la categoría *tres*, y el 6 con la *uno*.

El Gráfico 28 nos permite ver que:

- La dimensión ordena el número de empleadores que ha tenido el desempleado desde que terminó la carrera. En la parte negativa de dicha dimensión se sitúan las modalidades con menos empleados que en la parte positiva, que recoge las categorías con mayor movilidad de empleo.
- También nos ayuda a detectar la relación entre el clúster 3 y la categoría *dos*. La dimensión tres nos indica una relación entre el clúster 4 y la categoría *4 o más*.

Tabla 33. Resumen de las principales contribuciones a la inercia del número de empleadores y los clústeres de los perfiles laborales de desempleados

Dim.	Contribución a la inercia	Categoría	Signo coord.	Contribución a la inercia	Clúster	Signo coord.	Relación
DIM 1	0.4885	Tres (a)	Positivo	0.4151	5	Negativo	Inversa (a,5) Directa (a,6)
	0.4520	Uno (b)	Negativo	0.4926	6	Positivo	Directa (b, 5) Inversa (b,6)
DIM 2	0.2714	Uno (a)	Positivo	0.4605	5	Positivo	Directa (a,5) Directa(b,5) Inversa(c,5)
	0.2682	Tres (b)	Positivo		6	Positivo	Directa(a,6) Directa(b,6) Inversa(c,6)
	0.2433	Dos (c)	Negativo	0.307			
DIM 3	0.4621	4 o más(a)	Positivo	0.6866	4	Positivo	Directa (a,4) Inversa (b,4)
	0.4775	Dos (b)	Negativo				

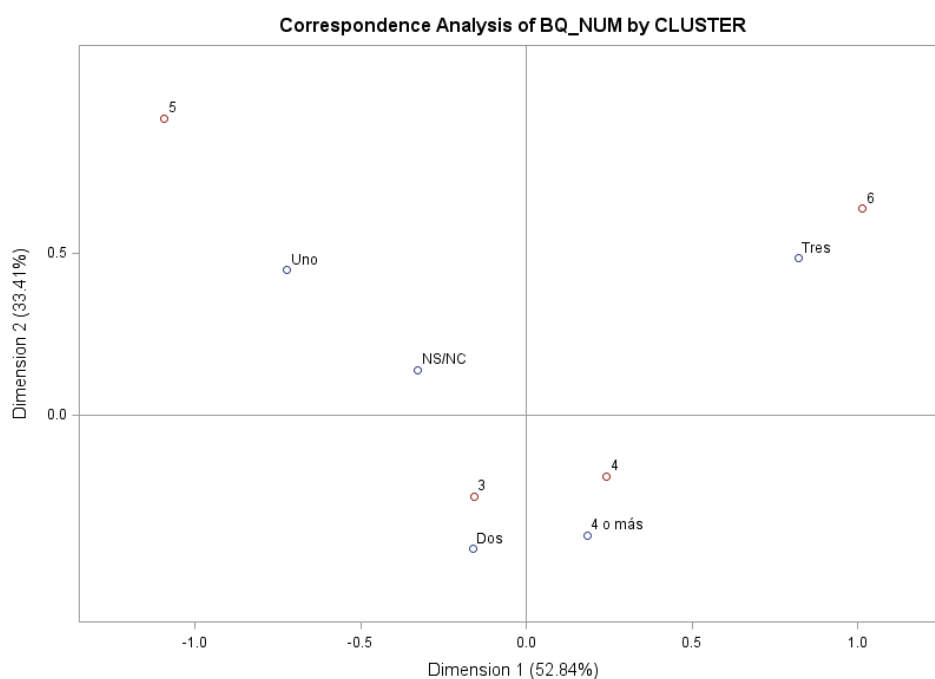


Gráfico 28. Representación de las modalidades sobre los ejes 1 y 2 del ACS del número de empleadores y los clústeres de perfiles laborales de individuos desempleados que sí han trabajado antes

- **Tiempo que ha estado trabajando desde que terminó los estudios frente a los clústeres**

El Gráfico 30 indica que las dos variables son independientes, y la inercia indica que debemos retener tres dimensiones con las que explicamos el 100% de la variabilidad de los datos.

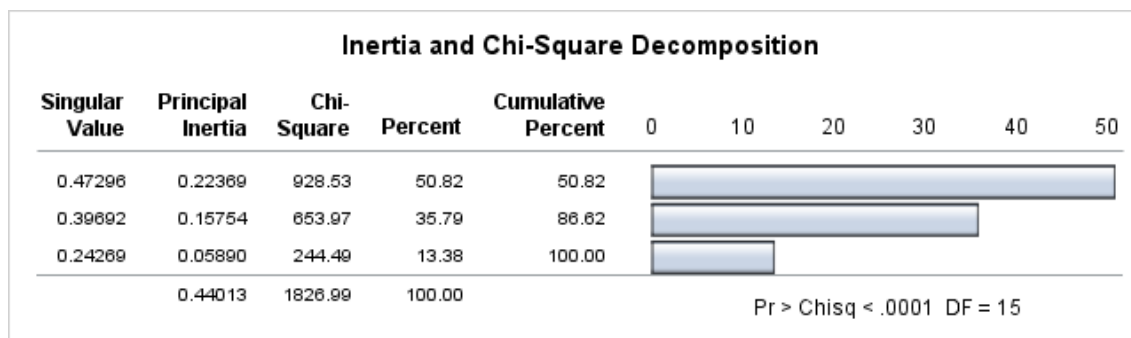


Gráfico 29. Descomposición de la inercia del ACS de la variable "tiempo que ha trabajado antes" de los perfiles laborales de individuos desempleados

La Tabla 34 muestra las categorías de cada variable que más contribuyen a cada dimensión. También nos apoyaremos en el Gráfico 31 de la dimensión uno y dos, que con ellas se explica el 86%.

Tabla 34. Resumen de las principales contribuciones a la inercia del tiempo que ha trabajado antes y los clústeres del perfil laboral de individuos desempleados

Dim.	Contribución a la inercia	Categoría	Signo coord.	Contribución a la inercia	Clúster	Signo coord.	Relación
DIM 1	0.8131	De 1 año a año y medio	Positivo	0.851	6	Positivo	Directa
DIM 2	0.3724	De 6 meses a 1 año (a)	Negativo	0.6538	5	Positivo	Inversa (a,5)
	0.5879	Menos de 6 meses (b)	Positivo				Directa (b,5)
DIM 3	0.5741	2 o más años (c)	Negativo	0.364	3	Negativo	Directa (c,3) Inversa(d,3)
	0.2389	De 1 año y medio a 2 años (d)	Positivo	0.4319	4	Positivo	Inversa (c,4) Directa (d,4)

La dimensión uno viene explicada mayoritariamente por la categoría de *De 1 año a año y medio* y se relaciona directamente con el clúster 6.

La dimensión dos está explicada por las dos categorías que recogen individuos que han estado menos de 1 año trabajando (*De 6 meses a 1 año y menos de 6 meses*). Aquellos

que han estado menos de 6 meses trabajando, se relacionan directamente con el clúster 5. Y los que han estado *de 6 meses a 1 año*, gráficamente observamos que se relacionan con el clúster 4.

En cuando a la dimensión dos está explicada por las categorías más altas, es decir, individuos que han estado más de 1 año y medio trabajando. Los que se sitúan en *2 o más años trabajando* están relacionados con el clúster 3, mientras que los que han estado de 1 año y medio a 2, con el 4.

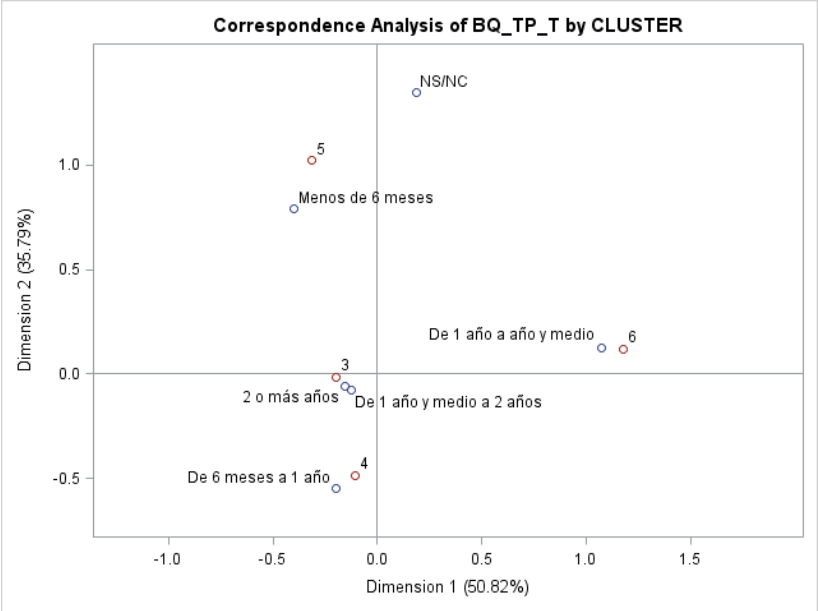


Gráfico 30. Representación de las modalidades sobre los ejes 1 y 2 del ACS del tiempo que ha trabajado antes y el perfil laboral de individuos desempleados

7.4.3. Caracterización de los individuos desempleados que no han trabajado nunca

Este grupo de individuos conforman un único clúster, dado que solo eran 1.342 individuos. Por ello, señalaremos la categoría mayoritaria de algunas de las variables que, tras ciertas pruebas, hemos considerado que ofrecen resultados de interés. Esta información se encuentra en la Tabla 35.

En el clúster de los individuos desempleados que no han trabajado nunca, el 60% de ellos son mujeres. Además, el grupo de edad más frecuente es el de menos de 30 años. Estos desempleados son de larga duración, ya que la mayoría lleva más de dos años en esta situación.

Tabla 35. Conjunto de variables utilizadas para caracterizar a individuos desempleados que no han trabajado nunca

Variable	Descripción	Categoría mayoritaria
SEXO	Sexo	Mujer
EDAD	Edad	Menores de 30 años
BSQEMP2	Momento en el que empezó a buscar trabajo	Después de acabar la carrera
BQTIEM2	Tiempo que lleva buscando empleo	2 o más años

BSQCO25	Búsqueda de empleo: preparando oposiciones	No se están preparando oposiciones
BSQCO21	Búsqueda de empleo: a través de internet	Busca el empleo en internet
BSQDIF3	Dificultades para encontrar empleo: falta de experiencia profesional	Muy importante

7.4.4. Caracterización de los individuos inactivos

Las variables con información más interesante en este grupo de individuos son las que recogemos en la Tabla 36. Donde, además, al igual que en el caso anterior, mostramos la categoría mayoritaria de cada variable para este grupo. Esta información se encuentra resumida en la siguiente tabla.

Tabla 36. Conjunto de variables utilizadas para caracterizar a individuos desempleados inactivos

Variable	Descripción	Categoría mayoritaria
SEXO	Sexo	Mujer
EDAD	Edad	Menores de 30 años
DISCA	Discapacidad	No
HA_TRAB	Ha tenido algún trabajo remunerado desde que finalizó la carrera	Sí

7.5. Resumen de la caracterización y descripción de los perfiles laborales

A continuación, mostramos la Tabla 37 donde resumimos las caracterizaciones hechas a lo largo de este punto para el clúster laboral, así como la frecuencia relativa de cada clúster.

A continuación, mostramos los perfiles obtenidos gracias al análisis anterior haciendo énfasis en sus características más diferenciadoras.

Perfil 1: Mujeres inactivas menores de 30 años.

Perfil 2: Desempleados de larga duración menores de 30 años que no han trabajado nunca.

Perfil 3: Hombres desempleados con experiencia previa y que han trabajado durante sus estudios.

Perfil 4: Mujeres desempleadas con experiencia previa entre uno y dos años, que han tenido entre 2 y 3 empleadores antes.

Perfil 5: Hombres trabajando autónomos relacionados con la industria manufacturera.

Perfil 6: Mujeres trabajando relacionadas con actividades sanitarias y servicios sociales con contrato temporal y dedicación a tiempo parcial.

Perfil 7: Hombres que trabajan no asalariados (autónomos, en prácticas o ayuda familiar) relacionados con la informática y comunicación.

Perfil 8: Individuos que trabajan asalariados tanto con contrato permanente como temporal.

Perfil 9: Mujeres trabajando en actividades relacionadas con la Administración Pública y Defensa y la Seguridad Social obligatoria asalariadas con contrato permanente y dedicación a tiempo completo y dentro del cuarto quintil de la base de cotización.

Tabla 37. Resumen de la caracterización del conjunto de perfiles laborales

Clúster	Situación profesional actual	Características principales	Frecuencia
CLUSTER 1	Inactivos	Mujeres, menores de 30 años	6.67
CLUSTER 2	Desempleados que no han trabajado nunca	Mujeres, menores de 30 años, desempleados de larga duración	4.65
CLUSTER 3	Desempleados que sí han trabajado	Mujeres, no han trabajado durante sus estudios, han trabajado 2 años o más, 2 empleadores	7.77
CLUSTER 4	Desempleados que sí han trabajado	Hombres, sí han trabajado durante sus estudios, han trabajado de 1 año y medio a 2 años, 4 o más empleadores	3.21
CLUSTER 5	Desempleados que sí han trabajado	Hombres, sí han trabajado durante sus estudios, han trabajado menos de 6 meses, 1 empleadores	1.42
CLUSTER 6	Desempleados que sí han trabajado	Mujeres, sí han trabajado durante sus estudios, han trabajado de 1 año a año y medio, 3 empleadores	1.97
CLUSTER 7	Trabajando	Mujeres, tiempo completo	21.93
CLUSTER 8	Trabajando	Hombres, Industria manufacturera, Empresario o trabajador independiente, tiempo completo	8.58
CLUSTER 9	Trabajando	Mujeres, Actividades sanitarias y de servicios sociales, contrato temporal, tiempo parcial	8.67
CLUSTER 10	Trabajando	Hombres, Quinto quintil, Información y comunicaciones, no asalariados, tiempo completo	4.76
CLUSTER 11	Trabajando	Hombres, primer quintil y quinto quintil, Información y comunicaciones, tiempo completo	5.85
CLUSTER 12	Trabajando	Contrato temporal	2.92
CLUSTER 13	Trabajando	Mujeres, segundo quintil, Asalariado con contrato permanente	5.03
CLUSTER 14	Trabajando	Mujeres, cuarto quintil, Administración Pública y defensa; Seguridad Social obligatoria, Asalariado con contrato permanente, tiempo completo	8.96
CLUSTER 15	Trabajando	Hombres, Asalariado con contrato permanente, tiempo completo	7.63

8. Relaciones entre los perfiles académicos y laborales

Una vez que tenemos tanto los clústeres del perfil académico como del laboral creados y caracterizados, vamos a tratar de encontrar la relación entre ambos. Para ello, llevaremos a cabo un análisis de correspondencias simples, a través del procedimiento descrito en la metodología.

El desarrollo de todos estos pasos se llevará a cabo mediante la información de las salidas ofrecidas por el procedimiento *proc corresp*.

```
proc corresp data=trabajo.union all print=both chi2p DIMENS=2;
tables CLUSLAB, CLUSAC;
run;
```

8.1. Estudio de las proporciones sobre el total de la población de cada modalidad

A continuación, analizamos la Tabla 38 de contingencia expresada en porcentaje, que se encuentra en el anexo. Debemos fijarnos en la última fila para el perfil académico, y en la última columna para el perfil laboral, donde vemos la presentación total de cada clúster.

En el perfil académico observamos dos modalidades que se encuentran por debajo del 5% de presencia que debería tener, y son el clúster 6 y 10. En cuanto a las modalidades del perfil laboral, son cuatro las que se encuentran por debajo de ese umbral, los clústeres 4, 5, 6 y 12.

Para saber si estas categorías con baja presencia sobre el total deben ser tratadas de manera especial, examinaremos las calidades de las mismas. Pero antes, debemos seleccionar el número de dimensiones con las que realizar el análisis. Sin entrar en mucho detalle en este punto, puesto que lo haremos más adelante, el número de ejes escogidos es dos. Ya que, además de rechazar en el contraste la hipótesis nula de independencia entre variables, con ellos logramos explicar el 96% de la variabilidad. A continuación, mostramos las calidades de la representación sobre los ejes de las modalidades de los clústeres “problemáticos”.

Tabla 38. Modalidades con una presencia inferior al 5% sobre el total de la población y su calidad de la representación sobre el eje

Modalidad	Variable	Calidad	Masa (frecuencia)
LAB_4	Clúster laboral	0.9170	0.0321
LAB_5	Clúster laboral	0.9315	0.0142
LAB_6	Clúster laboral	0.9971	0.0197
LAB_12	Clúster laboral	0.7730	0.0292
ACAD_6	Clúster académico	0.8258	0.0297
ACAD_10	Clúster académico	0.9869	0.0415

En ella vemos que todas las modalidades presentan una calidad superior al 70%, indicándonos que están bien representadas sobre los ejes por lo que continuamos el análisis con ellas.

8.2. Contraste de independencia de clústeres y aportaciones al Chi-cuadrado

Como adelantábamos en el punto anterior los perfiles académicos y laborales no son independientes. El valor del estadístico y la probabilidad de rechazar la hipótesis nula vienen dados en el Gráfico 31 ofrecido por el procedimiento *proc corresp*. El p-valor es inferior a 0.001 por lo que podemos rechazar la hipótesis nula.

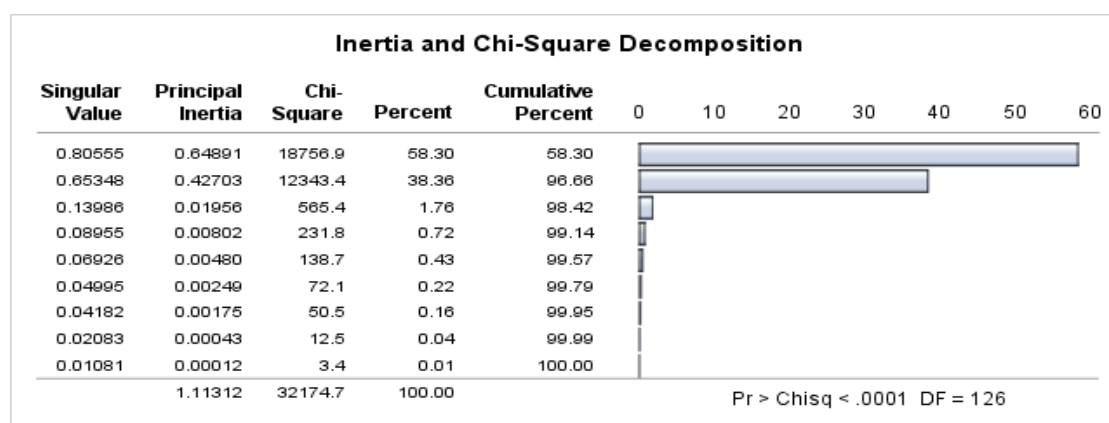


Gráfico 31. Descomposición de la inercia del ACS de clústeres de perfiles académicos y laborales

A continuación, vemos el mapa de calor (Gráfico 32) de las aportaciones al Chi-cuadrado de cada modalidad. Destacamos aquellos cruces que muestren un color más acentuado del azul. La más llamativa es el cruce entre el clúster *ACAD_5* y *LAB_3*, y nos indica que la relación entre ambas es fuerte.

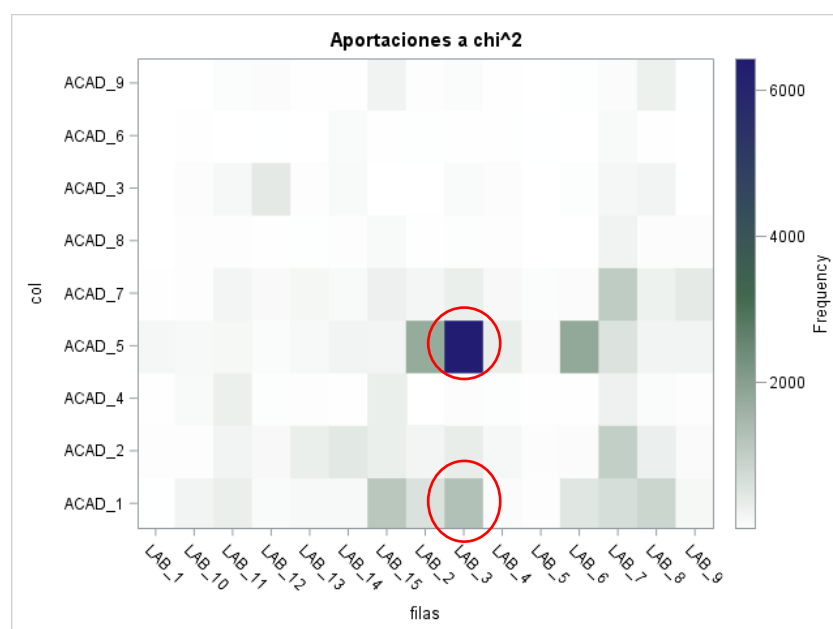


Gráfico 32. Mapa de calor de las aportaciones al Chi-Cuadrado de cada modalidad

Para saber el signo de dicha relación, observamos la Tabla 39, y nos fijamos en el signo de los residuo. Residuos por encima de los esperado bajo la hipótesis de independencia tendrán signo positivo e indicarán relaciones directas, mientras que residuos por debajo de lo esperado tendrán signo negativo y presentarán disociación. Vemos que el cruce de *ACAD_5* y *LAB_3* es de signo positivo, así que la relación será directa. Esto nos dice que una gran mayoría de los individuos del *ACAD_5* se encuentran en el *LAB_3*.

El clúster *ACAD_5* se trata de mujeres de ciencias sociales y jurídicas con titulaciones como diplomatura en Trabajo Social, Maestro-Especialidad Educación Infantil y licenciatura en Psicología, y el *LAB_3* son mujeres desempleadas que sí han trabajado alguna vez.

Un ejemplo de relación inversa, pero no tan intenso como el de antes, es el cruce de *ACAD_1* y *LAB_3*. En este caso, la presencia de los individuos del *ACAD_1* en el *LAB_3*, y viceversa, es minoritaria. En este caso, *ACAD_1* son estudiantes de ingeniería y arquitectura.

Tabla 39. Valores observados menos esperados de las modalidades de los clústeres de los perfiles académicos y laborales

Observed Minus Expected Values										
Perce nts	ACAD _1	ACAD _10	ACAD _2	ACAD _3	ACAD _4	ACAD _5	ACAD _6	ACAD _7	ACAD _8	ACAD _9
LAB_1	0,07	-0,09	0,45	0,04	-0,20	-0,63	0,01	0,27	0,10	0,00
LAB_10	0,73	-0,19	-0,32	0,35	0,46	-0,45	-0,10	-0,29	-0,23	0,04
LAB_11	1,03	-0,20	-0,90	0,58	0,86	-0,55	-0,04	-0,79	-0,24	0,25
LAB_12	0,34	-0,12	-0,49	0,71	0,23	-0,27	-0,06	-0,44	-0,11	0,21
LAB_13	-0,56	-0,15	0,99	-0,28	-0,32	-0,47	0,03	0,69	0,19	-0,11
LAB_14	-0,70	-0,24	1,59	-0,62	-0,30	-0,84	0,30	0,74	0,29	-0,22
LAB_15	2,28	-0,32	-1,23	0,12	1,02	-0,72	-0,14	-1,12	-0,47	0,57
LAB_2	-0,32	0,62	-0,77	0,02	-0,04	1,65	-0,11	-0,70	-0,16	-0,18
LAB_3	-0,97	1,12	-1,29	-0,54	-0,41	4,03	-0,14	-1,17	-0,29	-0,33
LAB_4	0,01	0,20	-0,53	0,28	0,24	0,62	-0,07	-0,48	-0,17	-0,09
LAB_5	0,09	0,10	-0,24	0,07	0,10	0,23	-0,04	-0,22	-0,04	-0,04
LAB_6	-0,32	0,35	-0,33	-0,18	-0,13	1,09	-0,03	-0,30	-0,06	-0,10
LAB_7	-2,89	-0,54	3,61	-1,21	-1,58	-2,06	0,51	3,54	1,16	-0,55
LAB_8	2,04	-0,34	-1,29	0,85	0,51	-0,81	-0,14	-1,14	-0,37	0,68
LAB_9	-0,82	-0,20	0,75	-0,19	-0,43	-0,82	0,03	1,42	0,40	-0,13

8.3. Análisis con las observaciones y modalidades elegidas

- Determinaremos el número de dimensiones a retener

A través de la figura que mostramos de la descomposición de la inercia, y el chi-cuadrado, determinaremos el número de dimensiones a retener. Nos basaremos en dos reglas para ello.

1. Los dos primeros factores ya recogen el 96%. Por lo que, según esta regla, con dos dimensiones será suficiente.
2. A través de la *average rule* deberíamos retener los factores cuya inercia superen 0.1236. En este caso, los dos primeros superan ese valor. Por lo que, según este criterio, debemos retener dos factores.

$$\frac{\text{Inercia}}{\min\{r-1, c-1\}} = \frac{1.11312}{\min\{14, 9\}} = 0.1236$$

Los dos criterios utilizados indican la elección de dos dimensiones para representar las variables.

- **Analizar la elección del número de dimensiones mediante un examen de la calidad**

Aunque en el punto anterior nos haya indicado la elección de dos dimensiones, debemos analizar que las modalidades de las dos variables en cuestión, quedan bien representadas sobre los ejes. Para ello, debemos comprobar que la calidad de cada categoría supera el 70%. Efectivamente, comprobamos a través la Tabla 40 ofrecida por el procedimiento SAS, en la primera columna *quality* que todas las modalidades superan el 70% establecido.

Tabla 40. Calidades de la representación sobre los ejes de los clústeres del perfil laboral

Summary Statistics for the Row Points	
	Quality
LAB_1	0.9102
LAB_10	0.9144
LAB_11	0.9631
LAB_12	0.7730
LAB_13	0.9742
LAB_14	0.9126
LAB_15	0.9246
LAB_2	0.9899
LAB_3	0.9965
LAB_4	0.9170
LAB_5	0.9315
LAB_6	0.9971
LAB_7	0.9982
LAB_8	0.9682
LAB_9	0.9215

Lo mismo ocurre con las modalidades que identifican los clústeres académicos, en todos los casos la calidad es superior al 70%, tal y como vemos en la Tabla 41.

Tabla 41. Calidades de la representación sobre los ejes de los clústeres del perfil académico

Summary Statistics for the Column Points	
	Quality
ACAD_1	0.9650
ACAD_10	0.9869
ACAD_2	0.9822
ACAD_3	0.7197
ACAD_4	0.9134
ACAD_5	0.9989
ACAD_6	0.8258
ACAD_7	0.9801
ACAD_8	0.9697
ACAD_9	0.8907

- **Análisis de los cosenos al cuadrado, y de la contribución parcial de cada modalidad**

En la Tabla 42 visualizamos los cosenos al cuadrado de las modalidades que conforman el perfil laboral en cada una de las dos dimensiones. En la tercera columna se recoge la suma de las dos dimensiones para cada modalidad que coincide con la calidad de la representación sobre los ejes.

Se puede observar en la tabla que la posición de los clústeres 1, 2, 3, 4, 5 y 6 en el plano bifactorial están principalmente determinada por la dimensión uno, los cuales coinciden con los inactivos y desempleados. Por el contrario, la posición de los clústeres 8, 10, 11, 12 y 15 está muy relacionado con la dimensión dos, y se corresponden con personas que se encuentran trabajando.

Tabla 42. Cosenos al cuadrado de los clústeres del perfil laboral

Squared Cosines for the Row Points			
	Dim1	Dim2	Quality
LAB_1	0.8999	0.0102	0.9101
LAB_10	0.0676	0.8469	0.9145
LAB_11	0.0042	0.9590	0.9632
LAB_12	0.0023	0.7707	0.773
LAB_13	0.5267	0.4475	0.9742
LAB_14	0.5529	0.3597	0.9126
LAB_15	0.0028	0.9219	0.9247
LAB_2	0.9733	0.0166	0.9899
LAB_3	0.9264	0.0701	0.9965
LAB_4	0.8389	0.0780	0.9169
LAB_5	0.8140	0.1175	0.9315
LAB_6	0.8989	0.0981	0.997
LAB_7	0.4871	0.5111	0.9982
LAB_8	0.0077	0.9605	0.9682
LAB_9	0.5784	0.3431	0.9215

En cuanto a los perfiles académicos, mostramos a continuación la Tabla 43 de los cosenos al cuadrado y su calidad de la representación. En este caso, la dimensión uno explica los clústeres 5 y 10, relacionados con la rama de ciencias sociales y jurídicas, y ciencias, respectivamente; y la dimensión dos, las modalidades 1, 3, 4 y 9. El clúster 1 y 4 están relacionados con la rama de ingeniería y arquitectura, mientras que el 9 con ciencias sociales y jurídicas.

Una vez hemos identificado la proporción de cada categoría explicada por cada dimensión, trataremos de buscar cuales son las modalidades que más explican a las dimensiones. Para ello, nos fijaremos en la contribución parcial a la inercia de cada categoría para cada dimensión. Buscamos que la contribución parcial acumulada supere el 70%.

Tabla 43. Cosenos al cuadrado de los clústeres del perfil académico

Squared Cosines for the Column Points			
	Dim1	Dim2	Quality
ACAD_1	0.0031	0.9619	0.9650
ACAD_10	0.8770	0.1098	0.9869
ACAD_2	0.4977	0.4845	0.9822
ACAD_3	0.0000	0.7197	0.7197
ACAD_4	0.0061	0.9073	0.9134
ACAD_5	0.9343	0.0646	0.9989
ACAD_6	0.4288	0.3970	0.8258
ACAD_7	0.4971	0.4831	0.9801
ACAD_8	0.3589	0.6107	0.9697
ACAD_9	0.0539	0.8367	0.8907

Comenzando con el perfil laboral, vemos en la Tabla 44 que los clústeres 2, 3 y 6 contribuyen en un 72% a la inercia del eje 1. Cabe destacar que estos tres clústeres también están mayoritariamente explicados por esta dimensión, como ya se vio en el análisis de los cosenos al cuadrado.

Por el contrario, la dimensión dos está explicada por los clústeres 7, 8, 11 y 15 que consiguen explicar el 70% de dicha dimensión. Al igual que en el caso anterior, estos clústeres están mayoritariamente explicados por la dimensión dos.

Tabla 44. Contribuciones parciales a la inercia de los clústeres del perfil laboral

Partial Contributions to Inertia for the Row Points		
	Dim1	Dim2
LAB_1	0.0138	0.0002
LAB_10	0.0025	0.0476
LAB_11	0.0003	0.1185
LAB_12	0.0001	0.0669
LAB_13	0.0260	0.0336
LAB_14	0.0382	0.0377
LAB_15	0.0004	0.2102
LAB_2	0.1523	0.0040
LAB_3	0.4331	0.0498
LAB_4	0.0417	0.0059
LAB_5	0.0149	0.0033
LAB_6	0.1285	0.0213
LAB_7	0.1132	0.1804
LAB_8	0.0010	0.1900
LAB_9	0.0339	0.0306

En cuanto al perfil académico (Tabla 45), las categorías que más contribuyen a la creación del eje son la 5 y la 10, alcanzando conjuntamente un 76%. Destacamos de nuevo, que estas dos categorías también se explican en la dimensión uno.

Por otro lado, la dimensión dos se explica por los clústeres 1, 2, 4 y 7. Cabe destacar que, en este caso, hacen falta más clústeres para alcanzar el umbral del 70%.

Tabla 45. Contribuciones parciales a la inercia de los clústeres del perfil académico

Partial Contributions to Inertia for the Column Points		
	Dim1	Dim2
ACAD_1	0.0006	0.2887
ACAD_10	0.1313	0.0250
ACAD_2	0.1071	0.1583
ACAD_3	0.0000	0.0870
ACAD_4	0.0005	0.1100
ACAD_5	0.6364	0.0669
ACAD_6	0.0088	0.0124
ACAD_7	0.0968	0.1430
ACAD_8	0.0157	0.0405
ACAD_9	0.0029	0.0683

A continuación, mostramos unas tablas a modo resumen de la información detallada anteriormente. En las columnas recogemos las categorías explicativas, y las modalidades con alta correlación con el eje, para cada dimensión, y, en las filas vemos las modalidades. Esta tabla la creamos tanto para el perfil académico (Tabla 46) como para el laboral (Tabla 47). Aparte de las conclusiones ya mencionadas previamente, en estas tablas se observan aquellos clústeres (*ACAD_6*, *ACAD_8*, *LAB_9*, *LAB_13* y *LAB_14*) que no se asocian con ninguna dimensión en especial.

Tabla 46. Resumen de los cosenos al cuadrado y las contribuciones parciales a la inercia de los clústeres del perfil académico

DIMENSIÓN 1			DIMENSIÓN 2	
Clústeres académicos	Categoría explicativa* ¹	Alta correlación con el eje * ²	Categoría explicativa* ¹	Alta correlación con el eje * ²
ACAD_1			X	X
ACAD_2			X	
ACAD_3				X
ACAD_4			X	X
ACAD_7			X	
ACAD_9				X
ACAD_5	X	X		
ACAD_10	X	X		
ACAD_6				
ACAD_8				

*¹Se consideran categorías explicativas aquellas con mayores contribuciones parciales hasta alcanzar una contribución parcial acumulada del 70%.

*² Se consideran una alta correlación con el eje cuando el coseno cuadrado es superior al 70%.

Tabla 47. Resumen de los cosenos al cuadrado y las contribuciones parciales a la inercia del perfil laboral

DIMENSIÓN 1		DIMENSIÓN 2		
Clústeres laborales	Categoría explicativa* ¹	Alta correlación con el eje * ²	Categoría explicativa* ¹	Alta correlación con el eje * ²
LAB_1		X		
LAB_2	X	X		
LAB_3	X	X		
LAB_4		X		
LAB_5		X		
LAB_6	X	X		
LAB_7			X	
LAB_8			X	X
LAB_9				
LAB_10				X
LAB_11			X	X
LAB_12				X
LAB_13				
LAB_14				
LAB_15			X	X

8.4. Representación gráfica y principales conclusiones

En el Gráfico 33 se representan las modalidades de ambos grupos de perfiles sobre las dos dimensiones retenidas. En el vemos que se dan tres agrupaciones principales, marcadas con un círculo rojo. Así mismo, se observa un clúster que no está en ninguno de estos cuatro grupos, por lo que lo analizaremos de manera aislada. A continuación, detallamos las principales conclusiones derivadas del ACS, teniendo en cuenta tanto el gráfico como las tablas anteriores.

En lugar de referirnos al nombre de los clústeres como tal, las conclusiones las obtendremos teniendo en cuenta las caracterizaciones realizadas anteriormente (Tabla 46 y Tabla 47).

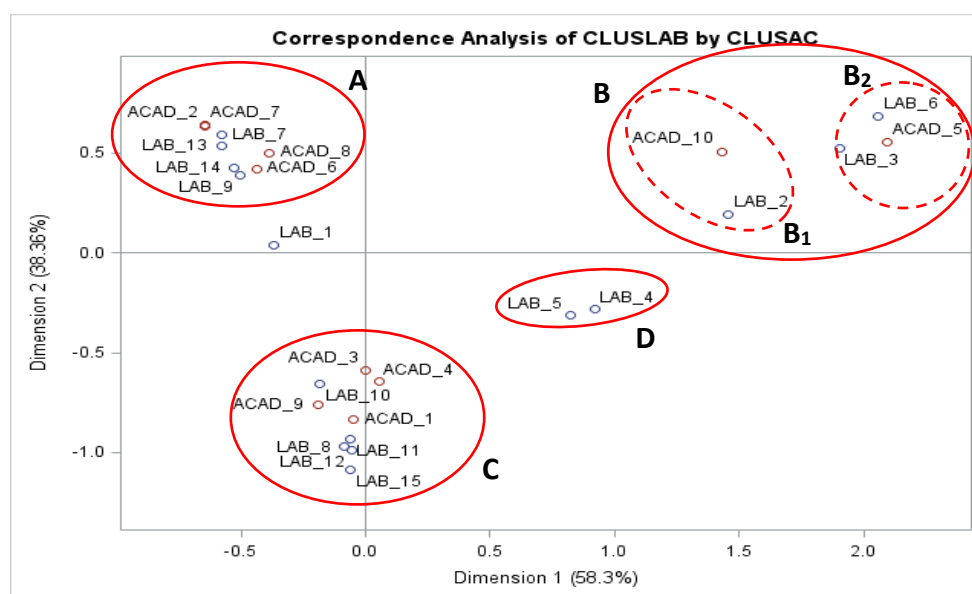


Gráfico 33. Representación sobre los ejes retenidos de los clústeres del perfil académico y laboral

Grupo A:

- El grupo A está formado únicamente por **mujeres** que se encuentran **trabajando**.
- Las ramas de estudio que conforman este grupo son: ciencias de la salud (Diplomatura en Enfermería, licenciatura en Medicina y diplomatura en Fisioterapia), arquitectura (Grado en Ingeniería de la Edificación, Arquitectura, Arquitectura Técnica), y artes y humanidades (Licenciatura en Filología Inglesa, licenciatura en Traducción e Interpretación y licenciatura en Historia).
- Destacamos que los egresados de ciencias de la salud no tienen ni idiomas acreditados, ni han cursado un máster, a diferencia de las otras dos ramas.
- Los egresados de arquitectura son mayoritariamente de universidades privadas.
- No observamos ningún patrón en cuanto a la temporalidad y tipo del contrato.

Grupo B:

El grupo B está formado por **mujeres desempleadas** que se pueden subdividir en los siguientes grupos:

- **Grupo B₁:**
 - Este grupo de desempleadas de **larga duración no ha trabajado previamente**.
 - Están relacionadas con la rama de ciencias, en concreto de ciencias naturales, con titulaciones como Biología, Química y Ciencias Ambientales.
 - Se trata de un grupo de personas muy formadas, ya que poseen un título de máster e idiomas acreditados.
- **Grupo B₂:**
 - Mujeres actualmente desempleadas con una **experiencia profesional superior a un año**, y entre **dos y tres empleadores diferentes**.
 - Las asociamos con titulaciones relacionadas con el área educativa, y las titulaciones: diplomatura en Trabajo Social, Maestro-Especialidad Educación Infantil y licenciatura en Psicología.
 - Se trata de egresadas que **no han realizado un máster y no tienen títulos de idiomas**.

Grupo C:

- **Hombres** que se encuentran **trabajando** actualmente a **tiempo completo**.
- Se les podría relacionar con el quinto quintil de base de cotización a la Seguridad Social.
- La rama de la titulación es ingeniería y ciencias sociales, en concreto, derecho, empresariales y economía.
- Mayoritariamente egresados de la universidad pública. Los ingenieros no poseen ni título de máster ni idiomas, mientras que los de ciencias sociales sí tienen idiomas acreditados.
- Especial mención merecen los grupos *ACAD_9* y *LAB_10*, que representan hombres autónomos, con salarios altos, de universidad privada, sin máster, pero con idiomas, en el área de la administración de empresas.

Grupo D:

- Grupo de **hombres desempleados** con experiencia previa, que han trabajado durante sus estudios.
- No se asocian con ningún perfil académico especial, pero dada su cercanía podríamos decir que se parecen entre sí.

LAB_1:

- Se trata de mujeres inactivas sin un perfil académico claro.

Teniendo en cuenta las conclusiones anteriores, podemos asignar una explicación a los ejes. La primera dimensión ordena las situaciones profesionales de individuos trabajando (valores negativos) a individuos desempleados (valores positivos). En cuanto a la segunda dimensión, divide los perfiles según su sexo: valores positivos están relacionados con mujeres, mientras que valores negativos se relacionan con hombres.

Nótese que cada uno de los cuatro grupos definidos anteriormente, se encuentran en un cuadrante diferente. No obstante, el grupo *LAB_1*, que es el más próximo al eje de coordenadas, tiene unas características particulares, y por eso lo observamos aislado.

9. Detección de factores influyentes que determinan la situación laboral

En este último apartado, vamos a detectar cuáles son los factores que más influyen a la hora de encontrar trabajo. Así mismo, para los individuos que se encuentren trabajando, trataremos de detectar los factores que más influyen en las características del mismo (i.e., duración del contrato o salario). Para ello haremos uso de los árboles de clasificación explicados en el apartado de metodología.

Utilizaremos variables relacionadas principalmente con aspectos académicos, junto con los clústeres de los perfiles académicos creados anteriormente y la experiencia laboral previa del individuo. Dadas las características de los árboles de clasificación, no será necesario hacer una selección de variables previa, ya que cuentan con un buen selector de variables. El conjunto de variables de entrada que utilizaremos para cada apartado lo recogemos en la Tabla 48.

Como podemos observar, todas las variables de la tabla han sido utilizadas previamente a excepción de la *experiencia*. Esta variable la hemos creado a partir de la variable que recoge información acerca del número de empleadores que ha tenido el individuo, tomando el valor 1 si este ha tenido experiencia laboral previa, y 0, en caso contrario. En cuanto al número de idiomas hablados, debido a que alguna de sus categorías tenía una frecuencia muy baja, la hemos reagrupado en menos categorías: 0, 1, 2 y 3 ó más.

Cabe destacar que en este apartado excluirémos a los individuos que sean inactivos debido a que tienen características particulares, tal y como dijimos en el apartado anterior.

Tabla 48. Variables utilizadas en los árboles de clasificación

Variable	Origen de la variable
Sexo	Encuesta
Edad	Encuesta
Tipo de Universidad	Encuesta
Rama de la titulación	Encuesta
Ha recibido beca	Encuesta
Tiene un máster	Encuesta
Rama del máster	Encuesta
Conocimientos informáticos	Encuesta
Ha realizado prácticas curriculares	Encuesta
Ha realizado prácticas extracurriculares	Encuesta
Número de idiomas hablados	Transformación
Idioma oficial	Nueva creación
Experiencia	Nueva creación
Clústeres del perfil académico	Nueva creación

9.1. Factores influyentes que determinan la presencia o ausencia de empleo

La variable objetivo en este caso es la situación laboral del individuo, que toma los valores: trabajando y desempleado.

En el Gráfico 34 se muestran las tasas de clasificación errónea del conjunto de *prueba* del proceso de *entrenamiento-prueba* repetido para los cuatro mejores modelos encontrados. Los diagramas de cajas de dicho gráfico permiten evaluar el comportamiento de los árboles tanto en media como en varianza.

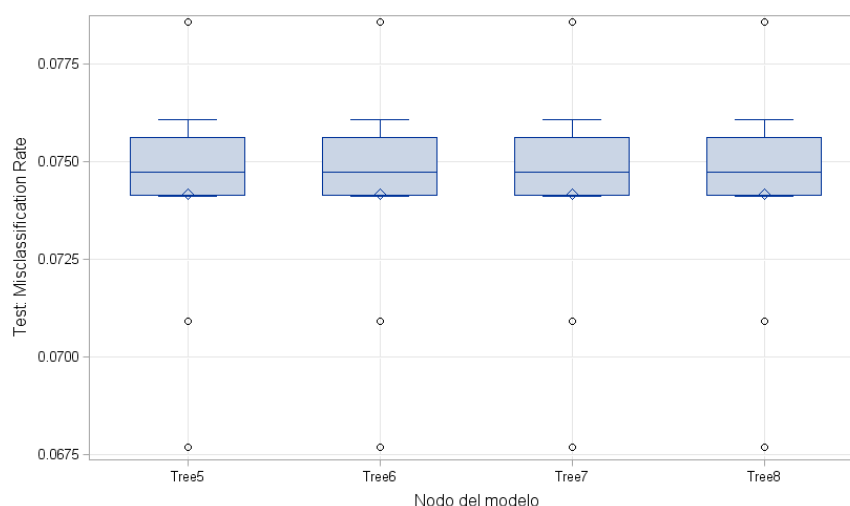


Gráfico 34. Diagrama de cajas de la tasa de clasificación errónea del árbol de clasificación de la situación laboral

En los diagramas de cajas observamos que no se aprecian grandes diferencias entre los valores del estadístico para cada uno de los cuatro árboles. Por lo tanto, al recurrir a la tabla de valores observamos que el árbol correspondiente al Chi-Cuadrado y con ajustes de profundidad es el que obtiene una menor tasa de error de clasificación en cada una de las diez repeticiones. La menor tasa clasificación errónea que obtiene es de 0.067688.

La Tabla 49 muestra la importancia de las variables utilizadas en la construcción del árbol (las variables que no aparecen no intervienen). Como podemos observar la variable más importante es el perfil académico que viene dado por los clústeres obtenidos anteriormente. Otras variables menos importantes son: rama, idioma acreditado oficialmente y si ha recibido beca durante los estudios.

Tabla 49. Importancia de variables del árbol de clasificación de la situación laboral

NAME	NRULES	IMPORTANCE
CLUSAC	1	1
RAMA	1	0,1421
IDIOMA_OFICIAL	1	0,116
EST_B1	1	0,1078

A continuación, mostramos el árbol (Gráfico 35) resultante en el que vemos los nodos creados. Los tonos de color oscuro observados en el árbol indican nodos más homogéneos mientras que los más claros hacen referencia a nodos con un reparto de trabajadores/desempleados más uniforme. Puesto que el objetivo de los árboles de clasificación es obtener nodos lo más homogéneos posible, nos centraremos en el análisis de aquellos nodos con colores más oscuros. Nótese que la distribución del nodo raíz es de un 79,64% de personas trabajando, y un 20,36% de desempleados.

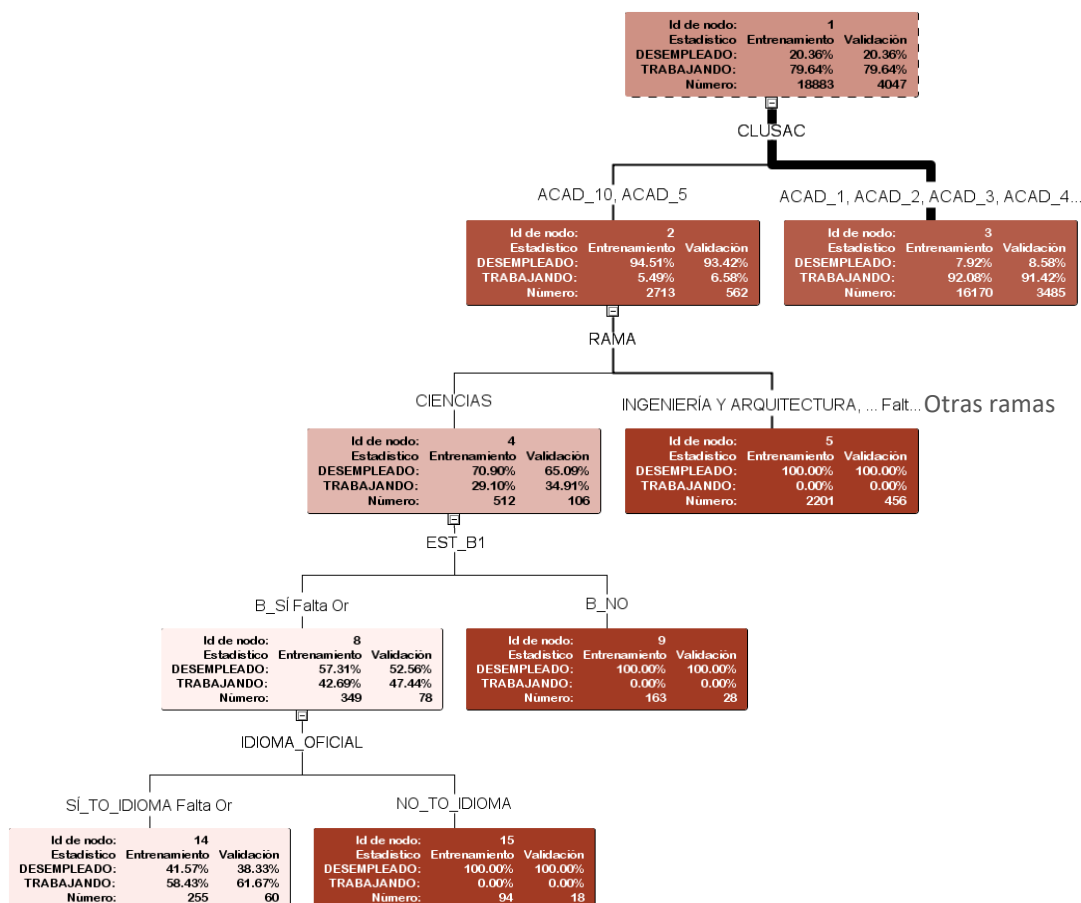


Gráfico 35. Diagrama del árbol de clasificación ganador de la situación laboral

Por lo tanto, siguiendo lo anterior, destacamos los siguientes aspectos:

- Siguiendo la rama de la derecha, vemos que todos los clústeres de los perfiles académicos obtenidos anteriormente, exceptuando los clústeres 5 y 10, están asociados a individuos trabajando. Más concretamente, el 91-92% de los individuos de estos clústeres se encuentran trabajando. Recordamos que la tabla (referencia) incluye un resumen de las características de los individuos de dichos clústeres.
- Siguiendo la rama de la izquierda, observamos que entre los clústeres 5 y 10 hay un 93-94% de desempleados. No obstante, si analizamos otras características de los individuos de estos clústeres conseguimos encontrar subgrupos de la población en los que el 100% están desempleados. Estos subgrupos se caracterizan por:
 - Individuos de ramas distintas a *ciencias*.
 - Individuos de la rama de *ciencias* que no han recibido beca durante sus estudios.
 - Individuos de la rama de *ciencias* que sí han recibido beca durante sus estudios, pero carecen de título oficial de idiomas.

9.2. Factores influyentes que determinan el tipo de contrato laboral

Una vez hemos determinado qué factores son los más influyentes a la hora de estar trabajando o desempleado, vamos a tratar de determinar cuáles son los que más influyen en el tipo de contrato. Para ello, solo tendremos en cuenta la población que se encuentra trabajando, y la variable objetivo tomará el valor 1 si el contrato es permanente, y 0 si es temporal.

En el Gráfico 36 vemos la tasa de clasificación errónea del conjunto de *prueba* del proceso de *entrenamiento-prueba* repetido. Como podemos observar, el mejor modelo tanto en media como en varianza es el *tree 11* que se ha obtenido utilizando el criterio del índice de Gini.

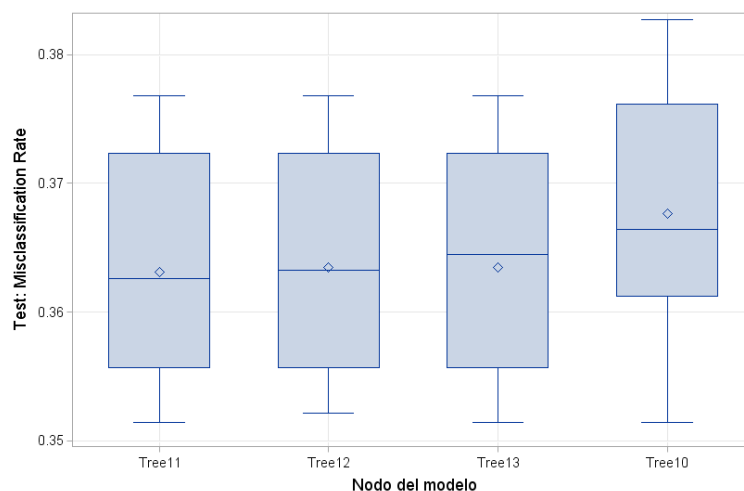


Gráfico 36. Diagrama de cajas de la tasa de clasificación errónea del árbol de clasificación de la situación profesional

Tabla 50. Importancia de variables del árbol de clasificación de la situación profesional

NAME	NRULES	IMPORTANCE
EDAD	3	1
CLUSAC	3	0,592
EXP	2	0,4564
TIC	2	0,301
RAMA	1	0,2082
EST2_MS	1	0,1756
T_UNIV	2	0,1739

Como hicimos con el árbol anterior, a continuación, mostramos en la Tabla 50 las variables más importantes para la construcción del mismo. En este caso, vemos que la variable más importante es la edad, seguidos de los clústeres de perfiles académicos y la experiencia previa laboral. Además, observamos que la importancia de las variables implicadas no es tan dispar, como en el caso anterior. El Gráfico 37 muestra el árbol obtenido y cómo podemos observar, se trata de un árbol con una profundidad mayor que el anterior.

En primer lugar, destacamos que el nodo raíz presenta muy poca homogeneidad puesto que aproximadamente el 53% de los individuos posee un contrato temporal. Siguiendo el procedimiento anterior, vamos a centrarnos en los nodos más homogéneos (representados con una tonalidad más oscura) para detectar los factores que influyen en el tipo de contrato, llegando a las siguientes conclusiones:

- Entre las personas de 35 años o más y que no tienen experiencia previa encontramos un 83% de asalariados con contrato permanente.
- Entre las personas con 35 años o más, experiencia previa y de la rama de *artes y humanidades* y *ciencias de la salud* encontramos un 71% de asalariados con contrato temporal.
- Entre los individuos con menos de 35 años pertenecientes a los clústeres de perfiles académicos 1, 4 y 9 de la rama de artes y humanidades, ciencias y ciencias de la salud, con nivel medio y bajo de uso de herramientas informáticas encontramos un 66% de asalariados con contrato temporal.
- Entre los individuos con edades entre 30 y 34 años, pertenecientes a los clústeres de perfiles académicos 2, 3 y 6 que estudiaron en una universidad privada encontramos un 63% de asalariados con contrato permanente.

En aras de la brevedad no comentaremos más nodos ya que presentan una menor homogeneidad. No obstante, se podría hacer un análisis similar a los anteriores.

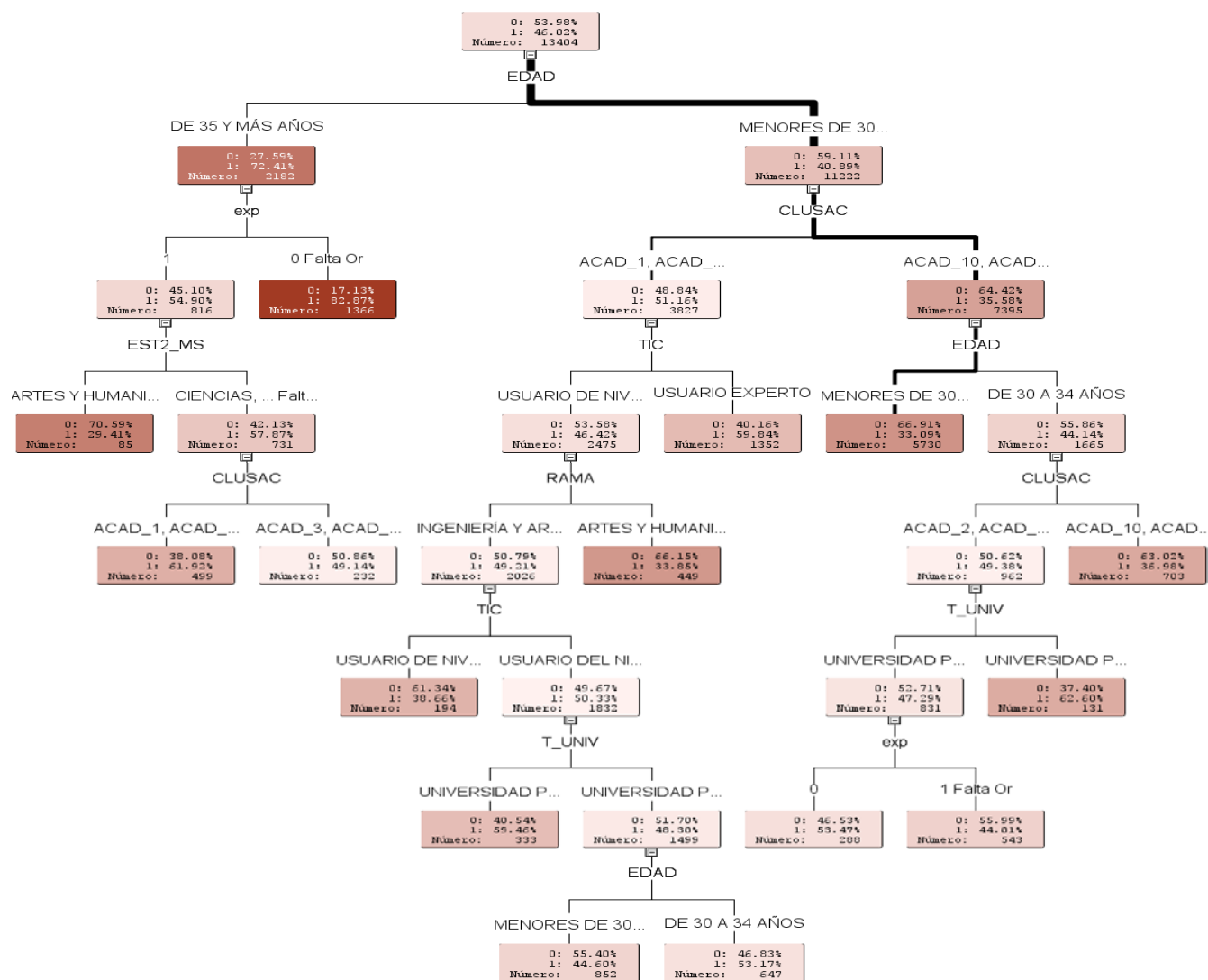


Gráfico 37. Diagrama del árbol de clasificación ganador de la situación profesional

10. Conclusiones

Para final este trabajo, a continuación, mostramos las principales conclusiones que podemos obtener gracias al análisis realizado:

1. Las universidades privadas tienen una mayor proporción de alumnos en titulaciones relacionadas con las ramas que presentan mayor potencial de empleo con respecto a las universidades públicas: ingeniería y arquitectura, derecho, empresariales y filología inglesa.
2. Los requisitos de los individuos para encontrar trabajando varían en función de la rama a la que pertenezca la titulación. Por ejemplo, los estudiantes de ingeniería y arquitectura que se encuentran trabajando no poseen ni título de máster ni idioma acreditado oficialmente, mientras que los de la rama de ciencias sociales sí poseen ambos títulos. También, estudiantes de ciencias naturales que se encuentran desempleados poseen tanto un título de máster como de idiomas.
3. Los individuos inactivos los podemos considerar como un grupo con características especiales y no podemos determinar un perfil claro de los mismos.
4. En general, no se han visto aspectos diferenciadores con respecto a la base de cotización a la Seguridad Social, a excepción de dos casos considerados como unos de los perfiles encontrados: hombres estudiantes en la universidad privada titulaciones como derecho y administración de empresas están relacionados con el quinto quintil de base de cotización; mujeres cuyo trabajo está relacionado con la Administración Pública y Defensa, y Seguridad Social Obligatoria se les asocia con el cuarto quintil.
5. No apreciamos que el sexo influya directamente en la situación laboral del individuo, pero sí observamos que existen ramas de conocimiento con una fuerte feminización/masculinización. Las ramas que asociamos con el desempleo están formadas mayoritariamente por mujeres, como es el caso de las ciencias naturales y las ciencias sociales (trabajo social, educación infantil y psicología). Por ello, podríamos decir que el sexo es un factor indirecto sobre la situación laboral del individuo, y la rama uno directo.
6. La situación profesional (contrato permanente o temporal) del individuo está determinada por la rama de estudio. Existen ramas en la que la temporalidad de los contratos es mayor, como en las artes y humanidad y ciencias de la salud.
7. Dentro del perfil de mujeres desempleadas hemos determinados tanto características académicas como laborales que definen el grupo, mientras que para el perfil de hombres desempleados no ha sido posible definirlos de forma tan clara.

11. Bibliografía

- [1] F. Rocha, «La crisis económica y sus efectos sobre el empleo en España», *Gac. Sind. Reflexión Debate*, vol. 19, pp. 67-90, 2012.
- [2] Á. Alonso Domínguez, «Análisis de la permanencia en el empleo de los trabajadores españoles durante el periodo 2007-2010», *Pap. Rev. Sociol.*, vol. 103, n.º 3, p. 339, abr. 2018.
- [3] «INEbase / Mercado laboral /Otras operaciones de mercado laboral /Encuesta de inserción laboral de titulados universitarios / Últimos datos». [En línea]. Disponible en:
https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176991&menu=ultiDatos&idp=1254735976597. [Accedido: 25-oct-2018].
- [4] «Instituto Nacional de Estadística. (Spanish Statistical Office)». [En línea]. Disponible en: <https://www.ine.es/>. [Accedido: 28-oct-2018].
- [5] J. de, «Encuesta de Inserción Laboral de los Titulados Universitarios EILU-2014», p. 45.
- [6] «SAS Procedures: SAS Procedures by Name :: SAS(R) Procedures by Name and Product». [En línea]. Disponible en:
<http://support.sas.com/documentation/cdl/en/allprodsproc/63875/HTML/default/viewer.htm#a003135046.htm>. [Accedido: 28-oct-2018].
- [7] C. Roever, N. Raabe, K. Luebke, U. Ligges, G. Szepannek, y M. Zentgraf, «Package 'klaR'». 19-mar-2018.
- [8] D. Peña, *Análisis de datos multivariantes*. Madrid: McGraw-Hill/Interamericana, 2002.
- [9] P. Jaccard, «Etude de la distribution florale dans une portion des Alpes et du Jura», *Bull. Soc. Vaudoise Sci. Nat.*, vol. 37, pp. 547-579, ene. 1901.
- [10] G. N. Lance y W. T. Williams, «A General Theory of Classificatory Sorting Strategies1. Hierarchical Systems», *Comput. J.*, vol. 9, n.º 4, pp. 373-380, feb. 1967.
- [11] «PROC CLUSTER: Clustering Methods :: SAS/STAT(R) 9.3 User's Guide». [En línea]. Disponible en:
https://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug_cluster_sect012.htm. [Accedido: 01-nov-2018].
- [12] A. Calviño, «Apuntes de la asignatura "COMPLEMENTOS DE FORMACIÓN EN TÉCNICAS DE MINERÍA DE DATOS"». Master en Minería de Datos e Inteligencia de Negocio, Universidad Complutense de Madrid.
- [13] «22540 - How can I tell how many clusters are in my data set?» [En línea]. Disponible en: <http://support.sas.com/kb/22/540.html>. [Accedido: 25-oct-2018].
- [14] Z. Huang, «Extensions to the k-means algorithm for clustering large data sets with categorical values», *Data Min. Knowl. Discov.*, vol. 2, n.º 3, pp. 283-304, 1998.
- [15] J. Llopis Pérez, «ANÁLISIS DE CORRESPONDENCIAS», *LA ESTADÍSTICA: UNA ORQUESTA HECHA INSTRUMENTO*, 06-jul-2013. .
- [16] T. J. Hastie, J. H. Friedman, y R. J. Tibshirani, *The elements of statistical learning: data mining, inference, and prediction*, 2nd. ed. corrected at 11th. print. New York: Springer, 2011.

A. Anexos

A.1. Tablas, gráficos y material complementario

1. Justificación de la selección de variables para el clúster académico.
 - 1.1. Pruebas para determinar las variables escogidas en el clúster de ámbito académico

Enumeramos a continuación todas las variables relacionadas con aspectos académicos, que encontramos en la metodología de la encuesta (referenciar). Se agrupan en las siguientes 12 categorías:

- Titulación realizada.
- Tipo de universidad y rama de conocimiento.
- Comunidad Autónoma de la Universidad en la que se ha titulado y ámbito de estudio.
- Discapacidad: Tiene discapacidad reconocida superior al 33%
- Prácticas curriculares y extracurriculares.

Variable	Descripción Variable
PRAC_CR	Realización de prácticas en empresas, instituciones o similares como parte del plan de estudios de la carrera
PRAC_CM	Meses en esas prácticas
EST_XCU	Realización de prácticas en empresas, instituciones o similares que no formasen parte del plan de estudios de la carrera
EST_X_M	Meses en esas prácticas

- Relación de becas.

Variable	Descripción Variable
EST_B21	Disfrutó de alguna beca: Beca general para el estudio
EST_B22	Disfrutó de alguna beca: Premio o beca de excelencia
EST_B23	Disfrutó de alguna beca: Beca de colaboración en la Universidad

EST_B24	Disfrutó de alguna beca: Beca de prácticas externas, para prestar un servicio o desempeñar un trabajo fuera de la Universidad.
EST_B25	Disfrutó de alguna beca: Beca para realizar estudios fuera de España (incluidas becas Erasmus)

- Estudios universitarios realizados en el extranjero y movilidad.

Variable	Descripción Variable
EST_M1	Realizó parte de sus estudios en el extranjero
EST_M2	Programa o beca de movilidad
EST_M3	País o región de la universidad de destino
EST_M4	Meses en el extranjero
EST_MES	Realizó parte de sus estudios en otra universidad española
EST2_TI	Tipo de otros estudios universitarios en el extranjero
EST2_FU	Otros estudios universitarios en el extranjero
EST2_TE	Momento en que se terminaron otros estudios universitarios en el extranjero

- Otra formación realizada (grado medio y grado superior).

Variable	Descripción Variable
EST2_2	Otros estudios: estudios artísticos superiores
EST2_3	Otros estudios: Ciclos Formativos o Enseñanzas Deportivas de grado superior
EST4	Estudios en curso fuera de los planes oficiales de estudio (cursos impartidos por academias, colegios profesionales, sindicatos, cursos en el centro de trabajo, cursos dirigidos a los desempleados, seminarios,

	conferencias, clases particulares recibidas, etc)
EST3	Estudios de grado medio de formación profesional, artes plásticas y diseño o enseñanzas deportivas completado

- Formación sanitaria especializada: Ha realizado o está realizando alguna formación sanitaria especializada (MIR, EIR, FIR, PIR, BIR, RFIR, QIR...etc)
- Realización de otros estudios oficiales en el momento de la entrevista.

Variable	Descripción Variable
EST2_MS	Rama de conocimiento del máster
EST2_T2	Momento en que se terminó el máster
EST2_NC	Estudios en curso

- Conocimiento de idiomas.

Variable	Descripción Variable
IDI_MT2	Idioma materno 2
IDIOMAS	Nº de idiomas que habla (sin contar los maternos)
IDIOMA	Nivel de idiomas
TITIDI1	Tiene título de la Escuela Oficial de Idiomas
TITIDI2	Tiene título en lengua extranjera reconocido oficialmente (PET, FCE, CAE, CPE, DELF, DALF, ...etc)
TITIDI3	Tiene título en lengua nacional reconocido oficialmente de cualquier nivel (A1, A2, B1, B2, C1, C2)

- Habilidades informáticas.

Las pruebas comenzarán seleccionando una variable representativa de cada grupo, en concreto, la que recoja más información a nivel general, sin entrar en demasiado detalle.

Realizaremos un análisis clúster jerárquico superficial con el objetivo de quedarse con tantos grupos como al menos consigamos explicar al menos el 70% de la variabilidad de los datos. Y a partir de ese umbral, buscaremos los máximos relativos explicados en la memoria.

En todos los casos, puesto que se trata de la muestra completa, y, tal y como dijimos en la memoria, el procedimiento *proc cluster* no funciona con conjuntos grandes de datos, hemos utilizado la matriz de distancias con las variables correspondientes en cada caso, sobre la muestra estratificada.

Una vez tengamos el número de grupos que consiga explicar dicho porcentaje, escogeremos el conjunto de variables ganador y realizaremos un análisis clúster en detalle en la memoria del presente trabajo. Mostraremos en este anexo las principales pruebas realizadas.

Para todas ellas, el método utilizado de cálculo de distancias entre clústeres será el de Ward. También, en todos los casos, mostraremos la tabla ofrecida por el procedimiento *proc cluster* con la información estadística del proceso de clusterización, además de un gráfico con los estadísticos *Pseudo F Statistics* y *Pesuto t-Squared*.

El conjunto de variables ganador es que el que consiga explicar mayor variabilidad con menos grupos, manteniendo siempre un mínimo de 6 variables, 3 de ellas fijas: sexo, rama y tipo de universidad.

- Prueba 1: 10 variables

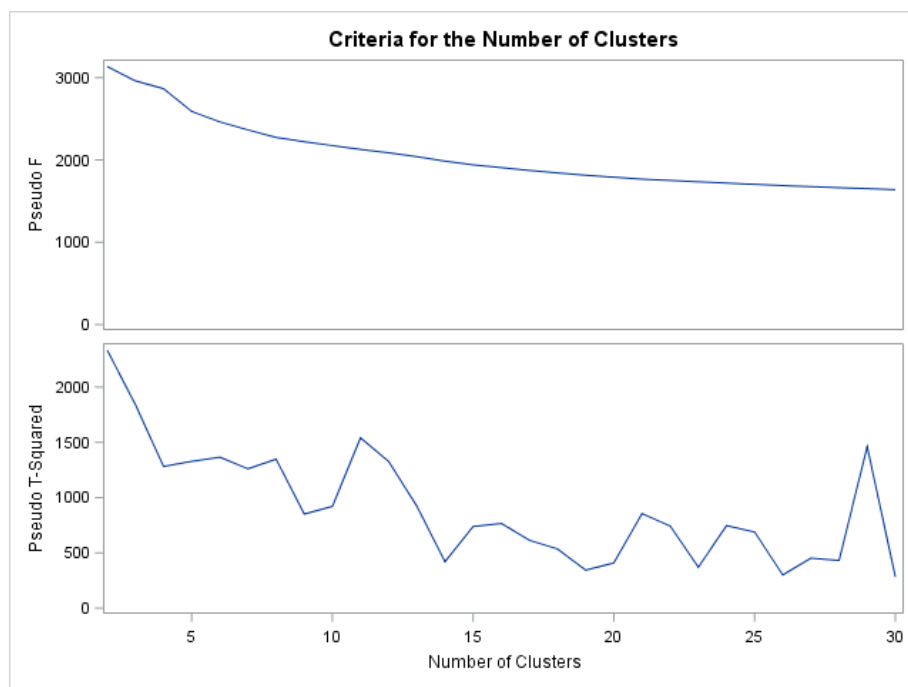
Variables a probar en esta primera prueba. Enumeraremos únicamente el nombre de las mismas, en el paso anterior se encuentra su descripción.

Resultado de la prueba: 30 clústeres necesarios para explicar el 70,4%.

SEXO	RAMA	T_UNIV	EST_B1	EST_MA	IDIOMA_OFICIAL	PRAC_CR	EST_M1	TIC	EST2_SA
------	------	--------	--------	--------	----------------	---------	--------	-----	---------

Cluster History								
Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square	Pseudo F Statistic	Pseudo t-Squared	Tie
30	CL49	CL77	915	0.0050	.704	1639	282	
29	CL189	CL60	1847	0.0056	.698	1652	1461	
28	CL51	CL36	1451	0.0064	.692	1663	432	
27	CL48	CL34	1300	0.0064	.686	1676	452	
26	CL40	CL42	1037	0.0069	.679	1689	301	
25	CL52	CL83	1029	0.0071	.672	1703	688	
24	CL53	CL69	1261	0.0073	.664	1720	746	
23	CL43	CL33	1307	0.0082	.656	1734	370	
22	CL50	CL35	1301	0.0083	.648	1751	743	
21	CL67	CL54	1155	0.0091	.639	1768	855	
20	CL58	CL31	1071	0.0091	.630	1789	409	
19	CL39	CL44	1041	0.0094	.621	1815	344	

18	CL28	CL47	1948	0.0099	.611	1843	535	
17	CL25	CL22	2330	0.0107	.600	1873	614	
16	CL38	CL32	1714	0.0113	.589	1906	767	
15	CL24	CL21	2416	0.0125	.576	1940	739	
14	CL26	CL20	2108	0.0126	.564	1985	421	
13	CL46	CL37	1489	0.0130	.551	2040	923	
12	CL29	CL27	3147	0.0159	.535	2088	1326	
11	CL62	CL41	1585	0.0190	.516	2128	1541	
10	CL11	CL16	3299	0.0209	.495	2175	921	
9	CL13	CL19	2530	0.0241	.471	2221	852	
8	CL12	CL23	4454	0.0274	.443	2273	1348	
7	CL17	CL30	3245	0.0280	.415	2365	1261	
6	CL15	CL8	6870	0.0340	.381	2462	1366	
5	CL10	CL7	6544	0.0398	.341	2590	1328	
4	CL18	CL9	4478	0.0406	.301	2866	1282	
3	CL4	CL14	6586	0.0722	.228	2961	1843	
2	CL3	CL6	13456	0.0929	.136	3135	2335	
1	CL2	CL5	20000	0.1355	.000	.	3135	

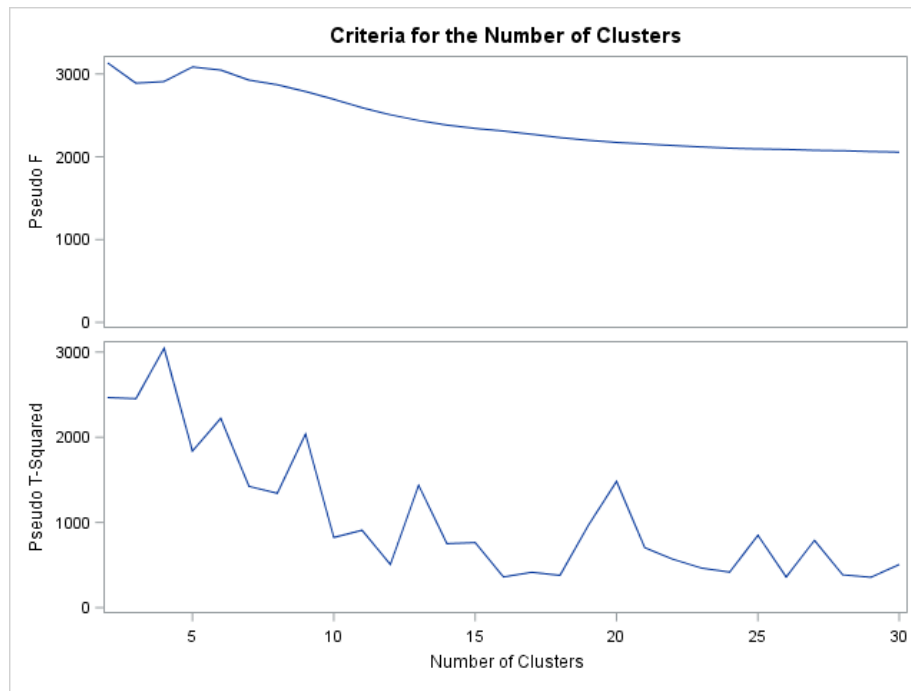


- Prueba 2: 9 variables

Resultado de la prueba: 26 grupos para explicar el 72,3%.

SEXO	RAMA	T_UNIV	EST_B1	EST_MA	IDIOMA_OFICIAL	PRAC_CR	EST_M1	TIC
------	------	--------	--------	--------	----------------	---------	--------	-----

Cluster History								
Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square	Pseudo F Statistic	Pseudo t-Squared	Tie
30	CL54	CL88	830	0.0055	.749	2056	507	
29	CL53	CL69	731	0.0059	.743	2064	357	
28	CL57	CL37	994	0.0061	.737	2073	385	
27	CL61	CL71	1190	0.0068	.730	2079	790	
26	CL31	CL42	1173	0.0068	.723	2089	359	
25	CL40	CL48	1562	0.0077	.716	2095	850	
24	CL45	CL49	736	0.0078	.708	2105	416	
23	CL46	CL51	991	0.0079	.700	2119	465	
22	CL38	CL50	1117	0.0081	.692	2137	567	
21	CL47	CL35	1412	0.0087	.683	2155	704	
20	CL52	CL68	1877	0.0093	.674	2174	1483	
19	CL33	CL55	1689	0.0095	.665	2199	964	
18	CL26	CL59	1465	0.0095	.655	2233	379	
17	CL23	CL44	1436	0.0097	.645	2273	415	
16	CL18	CL24	2201	0.0111	.634	2311	361	
15	CL22	CL43	1824	0.0130	.621	2342	763	
14	CL32	CL30	1764	0.0134	.608	2384	751	
13	CL20	CL56	2763	0.0138	.594	2438	1436	
12	CL29	CL17	2167	0.0143	.580	2507	507	
11	CL21	CL27	2602	0.0151	.565	2593	909	
10	CL34	CL28	2001	0.0166	.548	2694	824	
9	CL39	CL19	3116	0.0208	.527	2788	2038	
8	CL15	CL25	3386	0.0262	.501	2870	1342	
7	CL11	CL14	4366	0.0336	.468	2926	1424	
6	CL13	CL9	5879	0.0351	.433	3048	2222	
5	CL8	CL10	5387	0.0508	.382	3086	1838	
4	CL16	CL6	8080	0.0778	.304	2909	3047	
3	CL7	CL12	6533	0.0797	.224	2889	2453	
2	CL4	CL5	13467	0.0887	.135	3134	2467	
1	CL3	CL2	20000	0.1355	.000	.	3134	



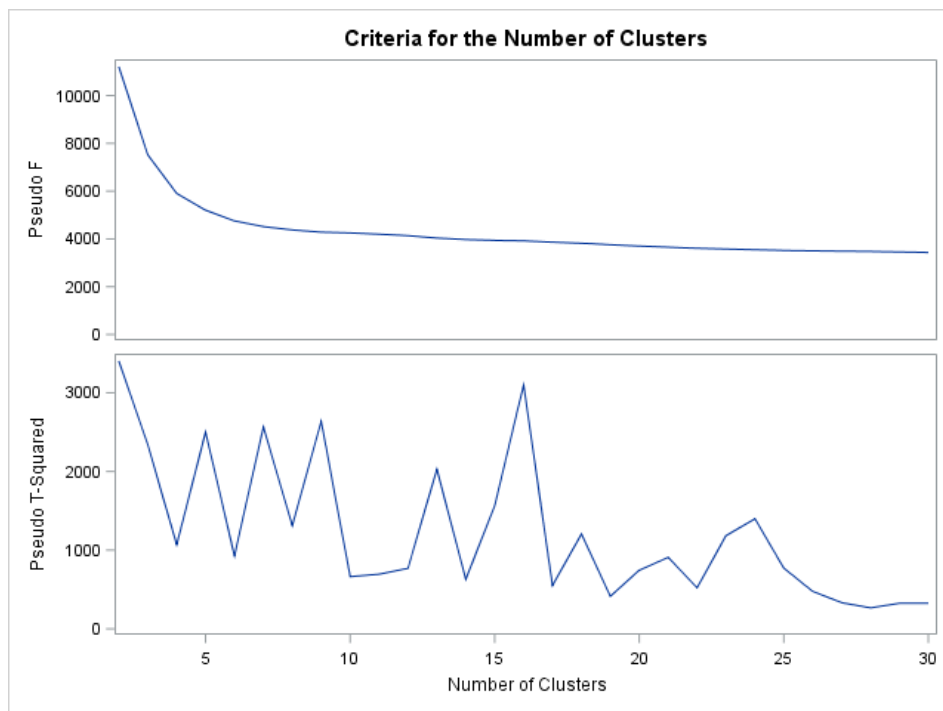
- Prueba 3: 8 variables

Resultado de la prueba: con 14 grupos conseguimos explicar el 72,1%.

SEX	RAMA_	T_UNI	EST2_	EST2_	IDIOMA_OFICI	EST_	TI
O	T	V	MA	MS	AL	M2	C

Cluster History								
Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square	Pseudo F Statistic	Pseudo t-Squared	Tie
30	CL77	CL82	583	0.0039	.833	3434	325	
29	CL33	CL58	1055	0.0040	.829	3456	324	
28	CL69	CL51	444	0.0044	.824	3475	267	
27	CL43	CL35	935	0.0050	.819	3487	333	
26	CL56	CL42	849	0.0053	.814	3501	476	
25	CL70	CL66	765	0.0055	.809	3519	771	
24	CL44	CL62	1786	0.0055	.803	3546	1398	
23	CL49	CL53	1406	0.0056	.798	3580	1185	
22	CL46	CL64	866	0.0061	.792	3613	522	
21	CL39	CL59	1125	0.0063	.785	3653	909	
20	CL23	CL65	1736	0.0066	.779	3700	745	
19	CL48	CL22	1240	0.0067	.772	3757	414	

18	CL61	CL32	1834	0.0073	.765	3820	1206	
17	CL37	CL25	1324	0.0089	.756	3865	550	
16	CL57	CL52	2363	0.0095	.746	3919	3099	
15	CL34	CL30	3220	0.0123	.734	3939	1565	
14	CL28	CL31	1045	0.0129	.721	3976	629	
13	CL24	CL54	2752	0.0131	.708	4038	2028	
12	CL29	CL38	1577	0.0134	.695	4133	769	
11	CL12	CL17	2901	0.0170	.678	4201	695	
10	CL11	CL14	3946	0.0206	.657	4255	663	
9	CL15	CL16	5583	0.0257	.631	4280	2638	
8	CL26	CL19	2089	0.0262	.605	4378	1311	
7	CL21	CL13	3877	0.0298	.575	4515	2566	
6	CL10	CL8	6035	0.0323	.543	4752	925	
5	CL9	CL18	7417	0.0331	.510	5202	2500	
4	CL6	CL27	6970	0.0402	.470	5904	1070	
3	CL7	CL20	5613	0.0404	.429	7522	2343	
2	CL5	CL3	13030	0.0701	.359	11E3	3404	
1	CL4	CL2	20000	0.3592	.000	.	11E3	

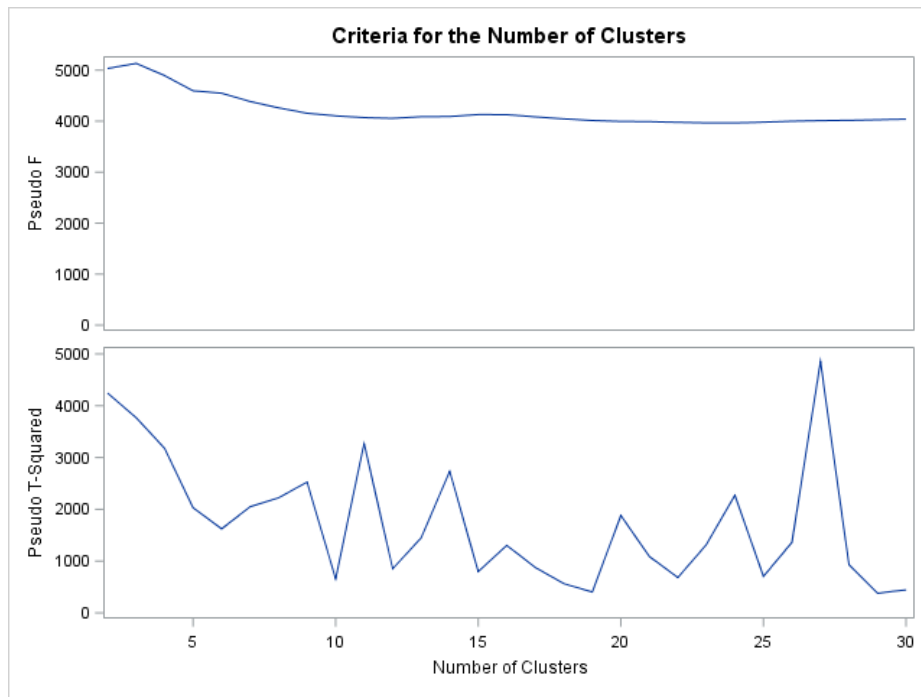


- Prueba 4: 7 variables

Resultado de la prueba: con 15 grupo conseguimos explicar el 74,3%

SEXO	RAMA	T_UNIV	EST2_MA	IDIOMA_OFICIAL	EST1_1	EST2_SA
------	------	--------	---------	----------------	--------	---------

Cluster History								
Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square	Pseudo F Statistic	Pseudo t-Squared	Tie
30	CL56	CL55	605	0.0048	.854	4039	444	
29	CL47	CL38	783	0.0048	.850	4027	380	
28	CL57	CL92	793	0.0050	.845	4018	929	
27	CL54	CL129	2332	0.0054	.839	4008	4858	
26	CL77	CL72	1036	0.0057	.833	3999	1362	
25	CL49	CL51	712	0.0064	.827	3981	704	
24	CL70	CL64	1360	0.0066	.820	3968	2269	
23	CL40	CL65	1421	0.0067	.814	3967	1320	
22	CL26	CL58	1316	0.0068	.807	3977	683	
21	CL31	CL36	2361	0.0072	.800	3991	1088	
20	CL50	CL45	1400	0.0082	.792	3994	1880	
19	CL43	CL32	749	0.0083	.783	4013	402	
18	CL22	CL48	1586	0.0084	.775	4047	565	
17	CL42	CL28	1278	0.0090	.766	4086	876	
16	CL41	CL44	1351	0.0100	.756	4125	1303	
15	CL30	CL37	1286	0.0128	.743	4129	797	
14	CL20	CL27	3732	0.0163	.727	4091	2732	
13	CL34	CL29	2304	0.0163	.711	4088	1451	
12	CL33	CL15	1950	0.0199	.691	4057	851	
11	CL52	CL24	2556	0.0200	.671	4070	3269	
10	CL19	CL12	2699	0.0217	.649	4105	658	
9	CL21	CL23	3782	0.0245	.624	4155	2525	
8	CL14	CL16	5083	0.0257	.599	4261	2223	
7	CL11	CL25	3268	0.0304	.568	4386	2050	
6	CL7	CL17	4546	0.0361	.532	4549	1622	
5	CL6	CL13	6850	0.0532	.479	4596	2030	
4	CL8	CL18	6669	0.0556	.423	4896	3179	
3	CL4	CL9	10451	0.0842	.339	5134	3772	
2	CL10	CL3	13150	0.1382	.201	5033	4245	
1	CL5	CL2	20000	0.2011	.000	.	5033	

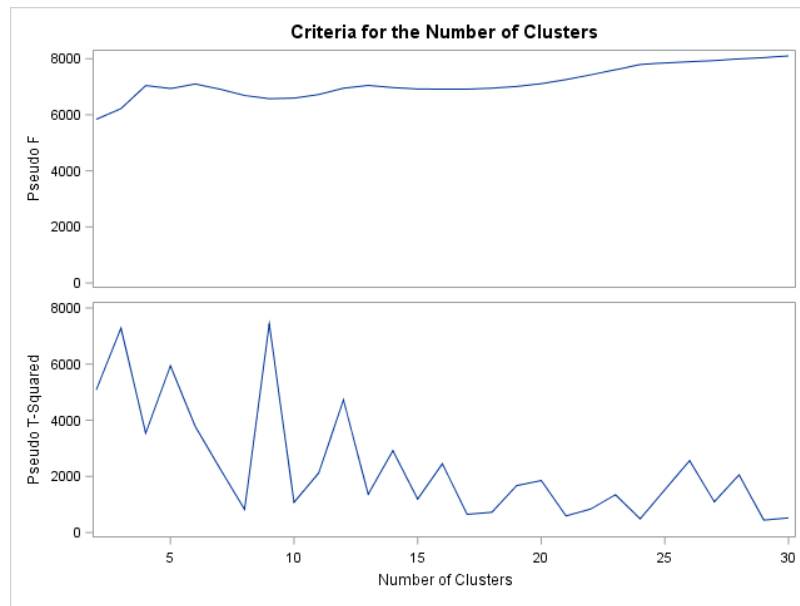


- Prueba 5: 6 variables

Resultado: con 6 grupos explicamos el 74,4%.

SEXO	RAMA	T_UNIV	EST2_MA	IDIOMA_OFICIAL	EST1_1	EST2_SA
------	------	--------	---------	----------------	--------	---------

Cluster History								
Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square	Pseudo F Statistic	Pseudo t-Squared	Tie
10	CL21	CL17	1868	0.0228	.748	6595	1076	
9	CL26	CL28	4092	0.0235	.725	6575	7428	
8	CL10	CL18	2763	0.0239	.701	6688	823	
7	CL15	CL16	3452	0.0259	.675	6915	2288	
6	CL12	CL14	4397	0.0351	.640	7100	3788	
5	CL9	CL13	6954	0.0586	.581	6935	5942	
4	CL7	CL11	5886	0.0674	.514	7043	3540	
3	CL5	CL6	11351	0.1303	.383	6219	7285	
2	CL8	CL3	14114	0.1574	.226	5840	5082	
1	CL4	CL2	20000	0.2260	.000	.	5840	



2. Determinar mejor método de cálculo de distancias para el clúster académico.

En este apartado encontramos las pruebas hechas para los métodos de enlace medio, del centroide y de Ward realizadas para el clúster académico. El objetivo era obtener el método que menor número de grupos explicase la mayor variabilidad de los datos posible.

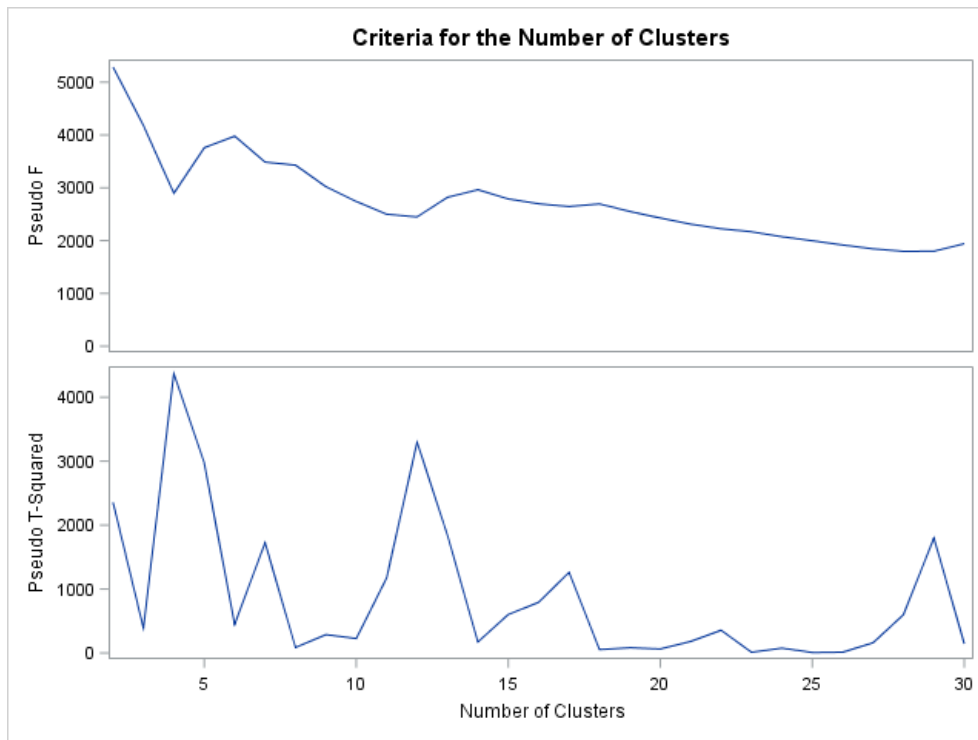
Mostraremos la tabla y gráfico de información estadística de la formación de los grupos ofrecida por el *proc cluster*. Buscaremos máximos relativos de los estadísticos.

- Método del enlace medio

Resultado de la prueba: 22 grupos explican el 70%

Cluster History									
Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square	Pseudo F Statistic	Pseudo t-Squared	Norm RMS Distance	Time
30	CL33	CL92	831	0.0020	.738	1942	144	0.6727	
29	CL38	CL72	4016	0.0219	.716	1802	1798	0.677	
28	CL39	CL94	2708	0.0081	.708	1796	601	0.6874	
27	CL46	CL177	1444	0.0022	.706	1845	160	0.6874	
26	CL47	CL103	50	0.0001	.706	1918	13.5	0.6897	

25	CL36	CL17 1	27	0.0001	.706	1997	7.4	0.6939	
24	CL31	CL40	135	0.0009	.705	2075	77.6	0.6958	
23	CL13 2	CL41	78	0.0002	.705	2168	12.1	0.6988	
22	CL56	CL10 7	1479	0.0044	.700	2224	356	0.6991	
21	CL54	CL23	1546	0.0020	.698	2313	181	0.725	
20	CL81	CL45	79	0.0007	.698	2427	63.4	0.7302	
19	CL10 0	CL25	332	0.0008	.697	2552	82.0	0.7526	
18	CL26	CL51	82	0.0007	.696	2693	54.4	0.7647	
17	CL43	CL55	2749	0.0168	.679	2646	1262	0.7752	
16	CL21	CL19	1878	0.0099	.669	2699	794	0.7808	
15	CL64	CL57	1079	0.0080	.661	2790	603	0.7911	
14	CL29	CL32	4100	0.0030	.658	2964	174	0.7935	
13	CL28	CL27	4152	0.0297	.629	2821		0.7964	
12	CL14	CL34	7155	0.0547	.574	2450	3298	0.799	
11	CL30	CL22	2310	0.0185	.556	2500	1176	0.8136	
10	CL35	CL44	381	0.0033	.552	2741	226	0.8325	
9	CL16	CL24	2013	0.0051	.547	3021	287	0.8693	
8	CL20	CL18	161	0.0017	.546	3430	86.5	0.9	
7	CL15	CL17	3828	0.0342	.511	3488	1727	0.9179	
6	CL7	CL10	4209	0.0126	.499	3980	447	0.9385	
5	CL13	CL11	6462	0.0696	.429	3760	2973	0.9683	
4	CL12	CL5	1361 7	0.1263	.303	2897	4365	0.9776	
3	CL9	CL8	2174	0.0082	.295	4178	394	1.0265	
2	CL4	CL3	1579 1	0.0856	.209	5286	2355	1.0419	
1	CL2	CL6	2000 0	0.2091	.000	.	5286	1.1657	

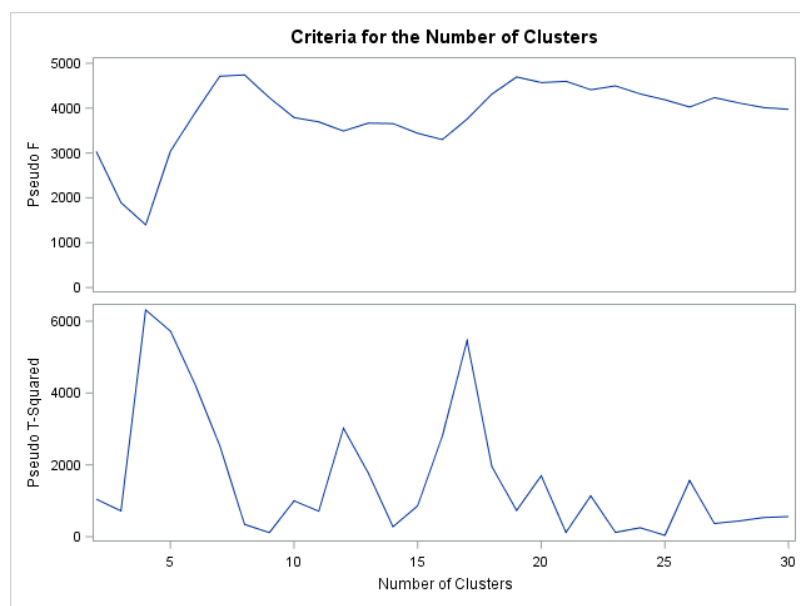


- Método del centroide

Resultado: 14 grupos explican el 70.4%

Cluster History									
Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square	Pseudo F Statistic	Pseudo T-Squared	Normalized Centroid Distance	Time
30	CL135	CL37	1324	0.0041	.852	3972	558	0.4735	
29	CL52	CL116	1226	0.0032	.849	4012	532	0.4794	
28	CL152	CL94	508	0.0014	.848	4114	437	0.4835	
27	CL100	CL121	301	0.0012	.846	4234	368	0.4847	
26	CL46	CL29	2048	0.0120	.834	4025	1564	0.4945	
25	CL90	CL45	49	0.0002	.834	4187	36.4	0.4977	
24	CL56	CL63	399	0.0017	.833	4317	250	0.5019	
23	CL59	CL58	93	0.0006	.832	4495	120	0.5033	
22	CL32	CL36	1945	0.0094	.823	4411	1138	0.5154	
21	CL23	CL31	159	0.0011	.822	4598	118	0.5245	

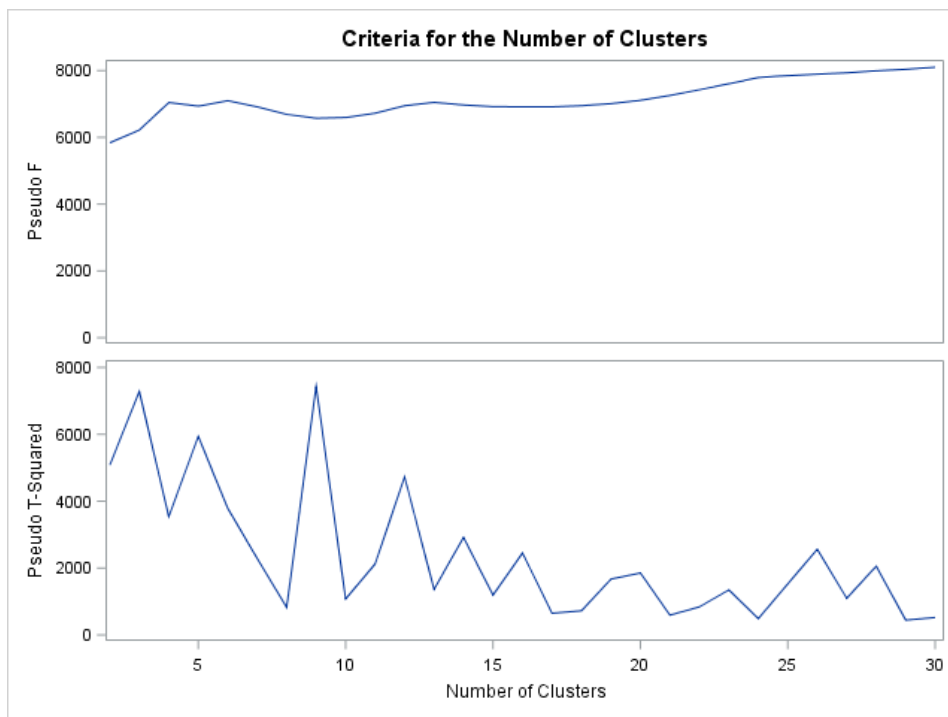
20	CL64	CL13 1	1213	0.0085	.813	4571	1698	0.532	
19	CL50	CL44	2482	0.0042	.809	4695	730	0.5385	
18	CL26	CL42	3324	0.0229	.786	4313	1955	0.5399	
17	CL19	CL35	4796	0.0354	.750	3756	5460	0.5435	
16	CL22	CL17	6741	0.0382	.712	3298	2804	0.5254	
15	CL28	CL27	809	0.0056	.707	3438	860	0.546	
14	CL24	CL10 5	493	0.0026	.704	3657	278	0.5857	
13	CL57	CL30	2067	0.0165	.688	3665	1780	0.588	
12	CL47	CL20	3758	0.0299	.658	3490	3021	0.6037	
11	CL15	CL53	1127	0.0088	.649	3693	709	0.6205	
10	CL16	CL34	7240	0.0183	.630	3790	998	0.6281	
9	CL21	CL25	208	0.0016	.629	4235	113	0.6527	
8	CL51	CL9	516	0.0049	.624	4739	343	0.6295	
7	CL13	CL38	3542	0.0382	.586	4713	2528	0.6658	
6	CL7	CL12	7300	0.0922	.494	3898	4238	0.711	
5	CL10	CL18	1056 4	0.1156	.378	3039	5726	0.7122	
4	CL6	CL5	1786 4	0.2049	.173	1396	6314	0.689	
3	CL8	CL14	1009	0.0139	.159	1894	717	0.7427	
2	CL3	CL11	2136	0.0275	.132	3035	1044	0.7184	
1	CL4	CL2	2000 0	0.1318	.000	.	3035	0.831	



- Método de Ward

Resultado: 10 grupos explican el 74,8%.

Cluster History								
Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square	Pseudo F Statistic	Pseudo t-Squared	Tie
30	CL60	CL34	770	0.0029	.922	8101	522	
29	CL83	CL66	425	0.0031	.919	8041	445	
28	CL44	CL230	1982	0.0032	.915	7994	2055	
27	CL63	CL92	672	0.0036	.912	7932	1096	
26	CL42	CL1926	2110	0.0036	.908	7894	2562	
25	CL206	CL205	1535	0.0039	.904	7851	.	
24	CL41	CL55	531	0.0044	.900	7793	488	
23	CL57	CL49	1143	0.0064	.893	7607	1348	
22	CL30	CL33	1327	0.0069	.886	7426	840	
21	CL29	CL36	747	0.0074	.879	7257	592	
20	CL56	CL54	1291	0.0078	.871	7111	1853	
19	CL37	CL46	1110	0.0079	.863	7010	1669	
18	CL27	CL38	895	0.0080	.855	6949	721	
17	CL32	CL24	1121	0.0083	.847	6913	651	
16	CL48	CL50	1420	0.0085	.839	6917	2452	
15	CL19	CL39	2032	0.0095	.829	6920	1193	
14	CL40	CL43	1667	0.0097	.819	6968	2918	
13	CL22	CL25	2862	0.0104	.809	7048	1366	
12	CL35	CL31	2730	0.0161	.793	6949	4730	
11	CL20	CL23	2434	0.0219	.771	6723	2124	
10	CL21	CL17	1868	0.0228	.748	6595	1076	
9	CL26	CL28	4092	0.0235	.725	6575	7428	
8	CL10	CL18	2763	0.0239	.701	6688	823	
7	CL15	CL16	3452	0.0259	.675	6915	2288	
6	CL12	CL14	4397	0.0351	.640	7100	3788	
5	CL9	CL13	6954	0.0586	.581	6935	5942	
4	CL7	CL11	5886	0.0674	.514	7043	3540	
3	CL5	CL6	11351	0.1303	.383	6219	7285	
2	CL8	CL3	14114	0.1574	.226	5840	5082	
1	CL4	CL2	20000	0.2260	.000	.	5840	



3. Tablas de contribuciones a la inercia y coordenadas, variable *rama* y *clúster* del clúster académico. Análisis de Correspondencias Simples.

- Rama:

Partial Contributions to Inertia for the Row Points			
	Dim1	Dim2	Dim3
Artes y humanidades	0.0063	0.6552	0.2320
Ciencias	0.0086	0.1429	0.7324
Ciencias de la salud	0.0445	0.0245	0.0031
Ciencias sociales y jurídicas	0.1777	0.1632	0.0318
Ingeniería y arquitectura	0.7628	0.0141	0.0006

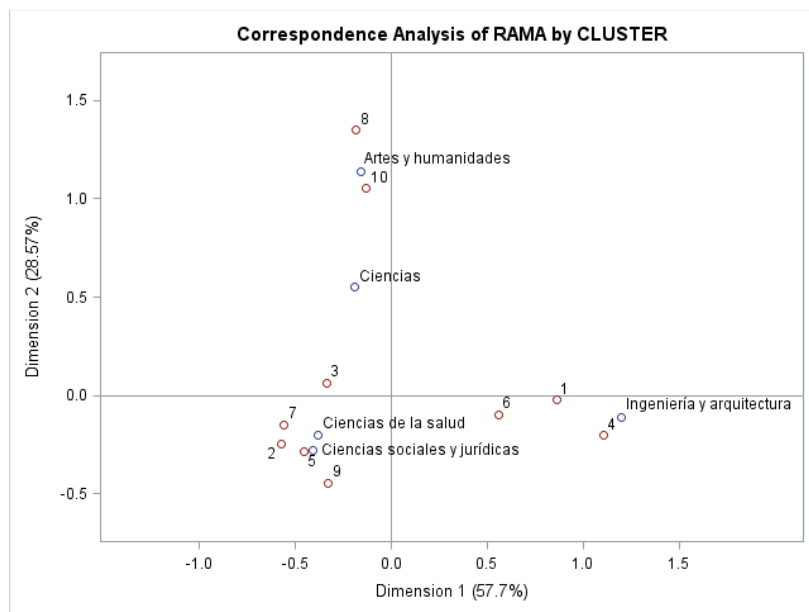
Row Coordinates			
	Dim1	Dim2	Dim3
Artes y humanidades	-0.1586	1.1361	-0.4123
Ciencias	-0.1927	0.5517	0.7617
Ciencias de la salud	-0.3818	-0.1991	0.0430
Ciencias sociales y jurídicas	-0.4102	-0.2766	-0.0744
Ingeniería y arquitectura	1.2021	-0.1152	-0.0150

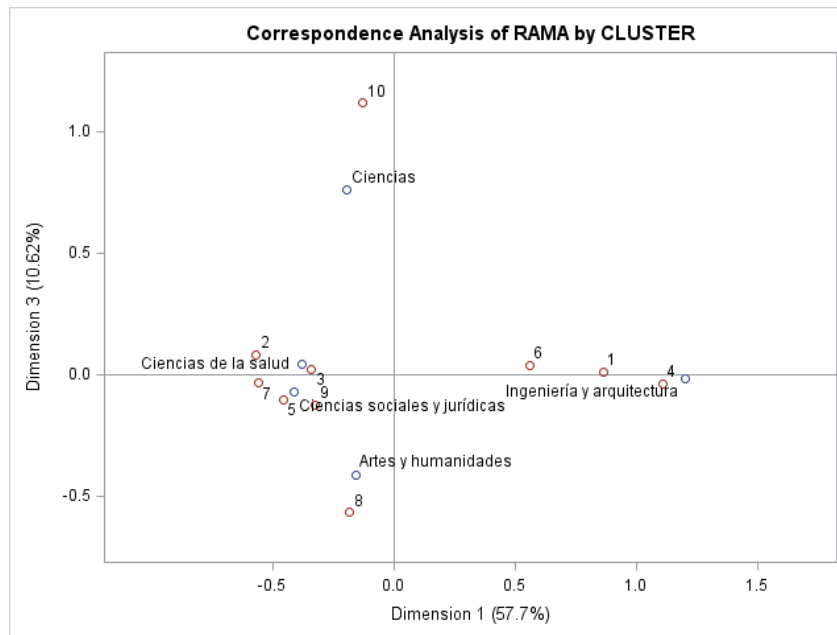
- Clúster del perfil académico: 1 al 10

Partial Contributions to Inertia for the Column Points			
	Dim1	Dim2	Dim3
1	0.3135	0.0005	0.0003
2	0.1289	0.0484	0.0138
3	0.0293	0.0021	0.0005
4	0.3264	0.0227	0.0023
5	0.0469	0.0363	0.0131
6	0.0220	0.0014	0.0005
7	0.1130	0.0169	0.0025
8	0.0054	0.6037	0.2861
9	0.0129	0.0480	0.0102
10	0.0017	0.2199	0.6707

Column Coordinates			
	Dim1	Dim2	Dim3
1	0.8626	-0.0249	0.0123
2	-0.5709	-0.2463	0.0801
3	-0.3392	0.0645	0.0187
4	1.1065	-0.2054	-0.0395
5	-0.4582	-0.2835	-0.1041
6	0.5587	-0.0976	0.0366
7	-0.5613	-0.1525	-0.0361
8	-0.1823	1.3519	-0.5675
9	-0.3268	-0.4444	-0.1247
10	-0.1298	1.0513	1.1195

- Representación de la dimensión tres frente a la una y a la dos.





4. Selección de variables del clúster de los individuos que están trabajando del clúster laboral.

El procedimiento será el mismo que para el clúster académico. En este caso, haremos las pruebas para el conjunto de individuos que se encuentran trabajando, y, por otro lado, los desempleados que si han trabajado antes.

En este caso, buscaremos explicar un 60% de la variabilidad de los datos.

3.1 Individuos trabajando

Todas las variables relacionadas con aspectos del trabajo actual son (referencia metodología de la encuesta):

- País o región en el que trabajaban en el momento de la entrevista.
- Situación profesional de su empleo actual.
- Tipo de jornada laboral en su empleo actual.
- Ocupación que desempeña en su empleo actual.
- Actividad económica principal de la empresa en la que trabaja actualmente.
- Tamaño de la empresa en la que trabaja actualmente.
- Nivel de formación más apropiado para desempeñar su trabajo actual.
- Área de estudios más apropiada para desempeñar su trabajo actual.

- Factores que han influido para conseguir su trabajo actual (sólo trabajadores por cuenta ajena).

Variable	Descripción
LTR_PAI	País en el que desempeña su labor profesional
SIT_PRO	Situación profesional actual
JORNADA	Tipo de jornada de trabajo actual
TRABOC	Ocupación que desempeña en el puesto de trabajo actual
TR_CNAE	Actividad principal del establecimiento o local donde trabaja actual
TR_TAM	Número de personas trabajando en el establecimiento o local donde trabaja actualmente
TR_ANIO	Año en el que empezó a trabajar en este último empleo
BCOTC11	Base de cotización a la SS en marzo de 2011 de los empleados por cuenta ajena (Quintiles)
BCOTC12	Base de cotización a la SS en marzo de 2012 de los empleados por cuenta ajena (Quintiles)
BCOTC13	Base de cotización a la SS en marzo de 2013 de los empleados por cuenta ajena (Quintiles)
BCOTC14	Base de cotización a la SS en marzo de 2014 de los empleados por cuenta ajena (Quintiles)

Categorías de la variable *TR_CNAE*:

TR_CNAE				
TR_CNAE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Actividades administrativas de oficina y otras actividades auxiliares a las empresas	419	1.45	419	1.45
Actividades artísticas, culturales, recreativas, deportivas y de entrenamiento	568	1.97	987	3.41
Actividades asociativas y sindicales	242	0.84	1229	4.25
Actividades de agencias viajes, operadores turísticos, servicios reservas y actividades relacionadas	141	0.49	1370	4.74
Actividades de fotografía	100	0.35	1470	5.09
Actividades de mantenimiento físico	68	0.24	1538	5.32
Actividades de organizaciones y organismos extraterritoriales	28	0.10	1566	5.42
Actividades de seguridad e investigación	88	0.30	1654	5.72
Actividades financieras y de seguros	581	2.01	2235	7.73
Actividades inmobiliarias	111	0.38	2346	8.12
Actividades profesionales, científicas y técnicas	2444	8.46	4790	16.57

Actividades relacionadas con el empleo	266	0.92	5056	17.49
Actividades sanitarias y de servicios sociales	3351	11.59	8407	29.08
Actividades veterinarias	174	0.60	8581	29.69
Administración Pública y defensa;Seguridad Social obligatoria	1444	5.00	10025	34.68
Agricultura, ganadería, silvicultura y pesca	253	0.88	10278	35.56
Comercio al por mayor y al por menor; reparación de vehículos de motor y motocicletas	1815	6.28	12093	41.84
Construcción e ingeniería civil	697	2.41	12790	44.25
Educación	3912	13.53	16702	57.78
Efectos personales y artículos de uso doméstico y otros servicios personales	83	0.29	16785	58.07
Hostelería	723	2.50	17508	60.57
Industria manufacturera	1236	4.28	18744	64.85
Industrias extractivas	55	0.19	18799	65.04
Información y comunicaciones	1445	5.00	20244	70.04
NS/NC	526	1.82	20770	71.86
No aplicable (TRAB_PRIN_1 = 2 ó 3)	7421	25.67	28191	97.53
Particulares como empleadores de personal doméstico;particulares:productores de bbss de uso propio	22	0.08	28213	97.61
Reparación de ordenadores	136	0.47	28349	98.08
Servicios a edificios y actividades de jardinería	65	0.22	28414	98.30
Suministro de agua, actividades de saneamiento, gestión de residuos y descontaminación	84	0.29	28498	98.59
Suministro de energía eléctrica, gas, vapor y aire acondicionado	78	0.27	28576	98.86
Transporte, almacenamiento, actividades postales y de correos	329	1.14	28905	100.00

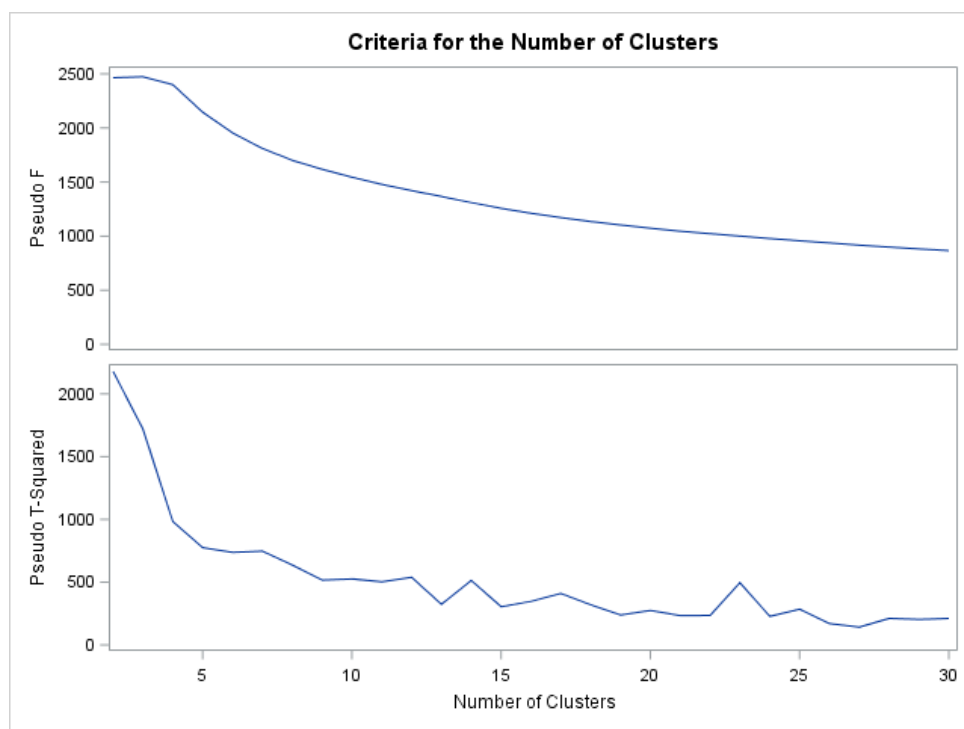
- Prueba 1: 9 variables

Resultado: con 30 grupos solo conseguimos explicar el 55,7%.

SEXO	SIT_PRO	JORNADA	TR_CNAE	BCOTC14	LTR_PAI	TRABOC	TR_TAM	TR_ANIO
------	---------	---------	---------	---------	---------	--------	--------	---------

Cluster History								
Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square	Pseudo F Statistic	Pseudo t-Squared	Tie
30	CL124	CL54	827	0.0038	.557	867	211	
29	CL39	CL55	1478	0.0044	.553	882	203	
28	CL99	CL101	739	0.0044	.549	899	212	

27	CL63	CL43	1070	0.0044	.544	917	142	
26	CL35	CL40	1098	0.0044	.540	937	170	
25	CL44	CL65	1188	0.0048	.535	957	285	
24	CL66	CL30	1319	0.0053	.530	978	228	
23	CL36	CL72	1356	0.0056	.524	1000	496	
22	CL51	CL28	1106	0.0061	.518	1022	235	
21	CL26	CL31	2178	0.0064	.512	1047	234	
20	CL61	CL45	1115	0.0065	.505	1073	275	
19	CL33	CL38	1327	0.0066	.499	1103	239	
18	CL37	CL50	2103	0.0071	.491	1136	320	
17	CL25	CL113	2006	0.0075	.484	1171	410	
16	CL24	CL78	2297	0.0075	.476	1212	347	
15	CL32	CL42	1570	0.0081	.468	1257	305	
14	CL23	CL52	1782	0.0082	.460	1310	513	
13	CL53	CL21	2761	0.0094	.451	1367	323	
12	CL29	CL41	2267	0.0118	.439	1421	538	
11	CL22	CL16	3403	0.0136	.425	1479	503	
10	CL34	CL12	2863	0.0150	.410	1545	526	
9	CL13	CL27	3831	0.0172	.393	1619	517	
8	CL19	CL20	2442	0.0199	.373	1701	637	
7	CL18	CL15	3673	0.0211	.352	1811	748	
6	CL10	CL11	6266	0.0238	.328	1955	737	
5	CL8	CL7	6115	0.0280	.300	2146	775	
4	CL5	CL14	7897	0.0355	.265	2401	983	
3	CL6	CL4	14163	0.0664	.198	2474	1720	
2	CL3	CL17	16169	0.0886	.110	2465	2179	
1	CL2	CL9	20000	0.1098	.000	.	2465	

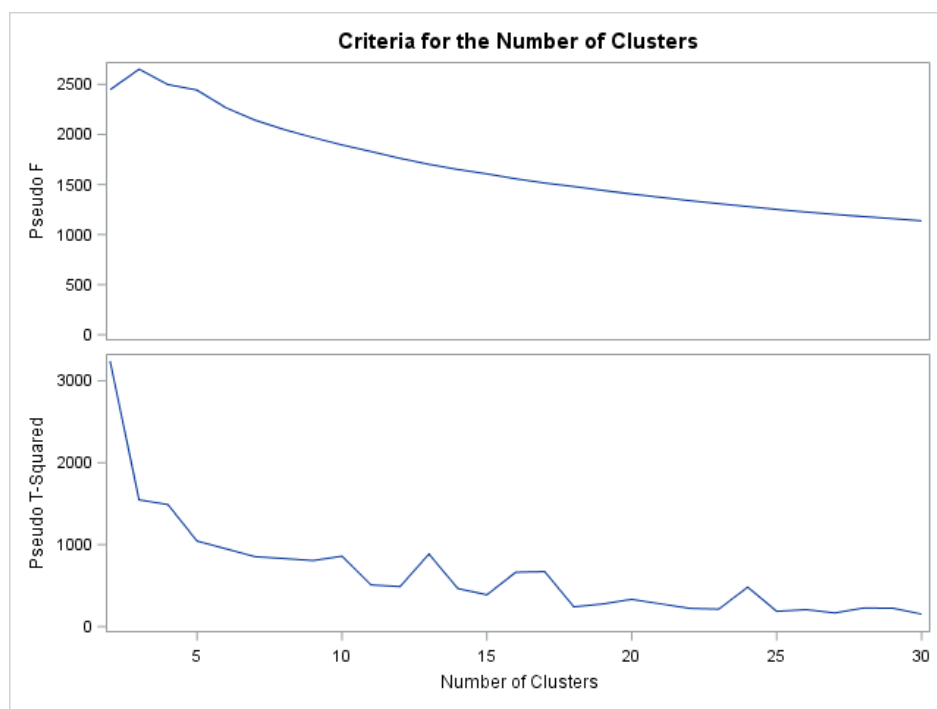


- Prueba 2: 8 variables

Resultado: con 25 grupos explicados el 60,5%

SEXO	SIT_PRO	JORNADA	TR_CNAE	BCOTC14	LTR_PAI	TRABOC	TR_TAM
------	---------	---------	---------	---------	---------	--------	--------

Cluster History								
Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square	Pseudo F Statistic	Pseudo t-Squared	Tie
30	CL47	CL128	847	0.0038	.623	1138	152	
29	CL36	CL75	982	0.0041	.619	1159	224	
28	CL29	CL92	1470	0.0044	.615	1179	227	
27	CL69	CL50	1094	0.0045	.610	1202	167	
26	CL41	CL61	807	0.0046	.605	1226	207	
25	CL32	CL70	1068	0.0048	.601	1252	186	
24	CL62	CL39	1226	0.0049	.596	1280	481	
23	CL76	CL30	1171	0.0054	.590	1309	213	
22	CL57	CL73	763	0.0057	.585	1339	222	
21	CL40	CL53	966	0.0061	.579	1371	277	
20	CL68	CL49	990	0.0066	.572	1405	333	
19	CL25	CL65	1587	0.0071	.565	1441	276	
18	CL35	CL26	1172	0.0075	.557	1479	241	
17	CL31	CL101	2005	0.0092	.548	1515	669	
16	CL24	CL34	1997	0.0093	.539	1557	664	
15	CL21	CL63	1677	0.0094	.529	1606	388	
14	CL23	CL28	2641	0.0118	.518	1649	463	
13	CL37	CL45	1920	0.0122	.505	1701	886	
12	CL18	CL48	2394	0.0130	.492	1762	487	
11	CL15	CL20	2667	0.0146	.478	1829	509	
10	CL22	CL33	2330	0.0174	.460	1895	859	
9	CL10	CL13	4250	0.0196	.441	1969	806	
8	CL27	CL43	2459	0.0230	.418	2049	832	
7	CL9	CL11	6917	0.0264	.391	2143	854	
6	CL12	CL19	3981	0.0297	.362	2266	949	
5	CL6	CL16	5978	0.0334	.328	2443	1044	
4	CL5	CL8	8437	0.0557	.273	2497	1488	
3	CL4	CL14	11078	0.0629	.210	2652	1546	
2	CL7	CL17	8922	0.1006	.109	2448	3241	
1	CL3	CL2	20000	0.1090	.000	.	2448	



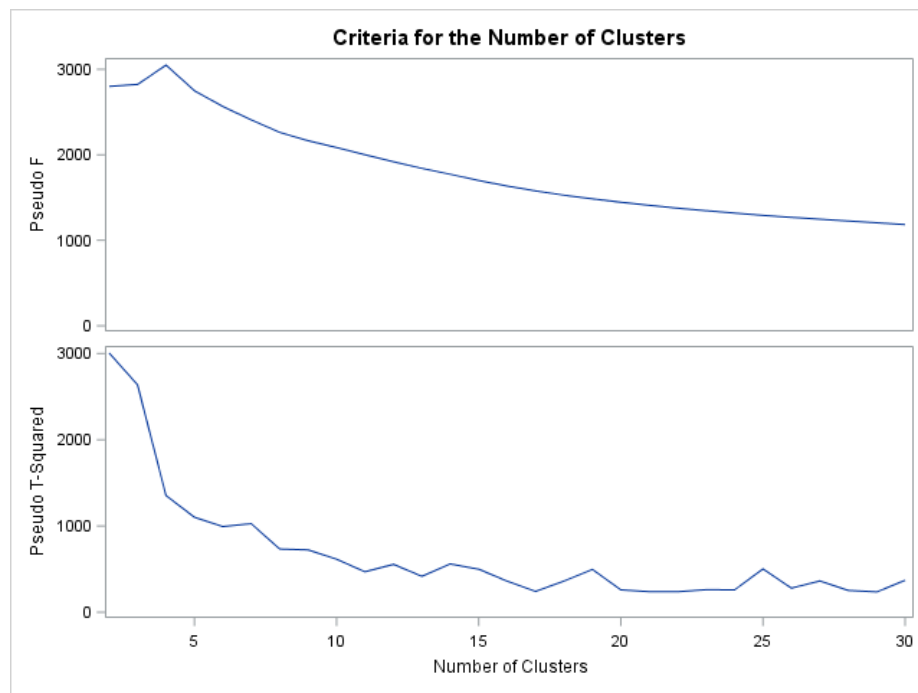
- Prueba 3: 7 variables

Resultado: con 26 grupos explicamos el 61,4%.

SEXO	SIT_PRO	JORNADA	TR_CNAE	BCOTC14	TRABOC	TR_TAM
------	---------	---------	---------	---------	--------	--------

Cluster History								
Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square	Pseudo F Statistic	Pseudo t-Squared	Tie
30	CL78	CL114	859	0.0038	.632	1184	371	
29	CL62	CL92	896	0.0041	.628	1205	236	
28	CL41	CL45	1392	0.0045	.624	1226	253	
27	CL43	CL67	1219	0.0049	.619	1247	363	
26	CL36	CL53	1219	0.0051	.614	1269	280	
25	CL65	CL47	1335	0.0054	.608	1292	503	
24	CL44	CL109	722	0.0055	.603	1318	259	
23	CL29	CL37	1531	0.0056	.597	1346	262	
22	CL55	CL31	1141	0.0059	.591	1376	238	
21	CL38	CL39	1403	0.0060	.585	1409	239	
20	CL24	CL68	1337	0.0063	.579	1446	259	
19	CL42	CL54	1902	0.0068	.572	1484	497	
18	CL28	CL64	2011	0.0069	.565	1528	362	
17	CL52	CL22	1516	0.0070	.558	1578	242	

16	CL70	CL26	1606	0.0073	.551	1634	361	
15	CL30	CL34	1612	0.0074	.543	1699	499	
14	CL27	CL121	2005	0.0080	.535	1772	561	
13	CL33	CL16	2603	0.0103	.525	1842	416	
12	CL23	CL19	3433	0.0115	.514	1919	556	
11	CL35	CL13	3274	0.0133	.500	2001	470	
10	CL17	CL48	2610	0.0160	.484	2086	615	
9	CL10	CL18	4621	0.0202	.464	2164	722	
8	CL32	CL21	2383	0.0220	.442	2263	733	
7	CL25	CL20	2672	0.0225	.420	2408	1026	
6	CL9	CL15	6233	0.0288	.391	2565	993	
5	CL6	CL7	8905	0.0359	.355	2749	1101	
4	CL12	CL8	5816	0.0409	.314	3049	1353	
3	CL5	CL11	12179	0.0937	.220	2823	2636	
2	CL4	CL14	7821	0.0974	.123	2800	3006	
1	CL2	CL3	20000	0.1228	.000	.	2800	



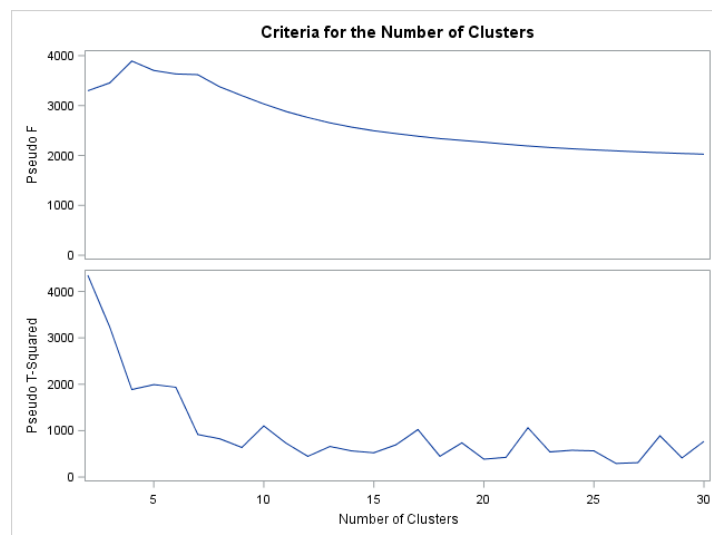
- Prueba 4: 6 variables

Resultado: con 13 grupos explicamos el 61,4%

SEXO	SIT_PRO	JORNADA	TR_CNAE	BCOTC14	TR_TAM
------	---------	---------	---------	---------	--------

Cluster History

Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square	Pseudo F Statistic	Pseudo t-Squared	Tie
30	CL43	CL95	1904	0.0048	.746	2024	769	
29	CL106	CL86	514	0.0052	.741	2040	412	
28	CL66	CL39	1491	0.0057	.735	2055	890	
27	CL32	CL112	1521	0.0057	.730	2072	311	
26	CL81	CL49	861	0.0058	.724	2093	293	
25	CL52	CL53	948	0.0064	.717	2112	564	
24	CL70	CL44	1199	0.0066	.711	2133	577	
23	CL65	CL33	1410	0.0068	.704	2159	542	
22	CL123	CL71	1101	0.0068	.697	2189	1062	
21	CL34	CL69	1482	0.0069	.690	2225	424	
20	CL27	CL46	2109	0.0072	.683	2266	385	
19	CL67	CL31	1691	0.0084	.675	2301	738	
18	CL47	CL29	897	0.0090	.666	2339	448	
17	CL22	CL100	2005	0.0092	.656	2385	1023	
16	CL25	CL36	2250	0.0097	.647	2438	695	
15	CL41	CL45	1076	0.0105	.636	2495	524	
14	CL19	CL37	2292	0.0106	.625	2568	564	
13	CL38	CL23	1948	0.0112	.614	2653	659	
12	CL18	CL40	1383	0.0112	.603	2761	446	
11	CL13	CL28	3439	0.0127	.590	2882	734	
10	CL24	CL30	3103	0.0132	.577	3032	1102	
9	CL15	CL20	3185	0.0157	.561	3200	638	
8	CL16	CL12	3633	0.0198	.542	3376	824	
7	CL26	CL21	2343	0.0211	.521	3619	916	
6	CL10	CL7	5446	0.0446	.476	3632	1936	
5	CL11	CL8	7072	0.0504	.426	3704	1993	
4	CL5	CL14	9364	0.0570	.369	3891	1887	
3	CL4	CL9	12549	0.1118	.257	3455	3232	
2	CL6	CL17	7451	0.1153	.142	3297	4349	
1	CL2	CL3	20000	0.1415	.000	.	3297	

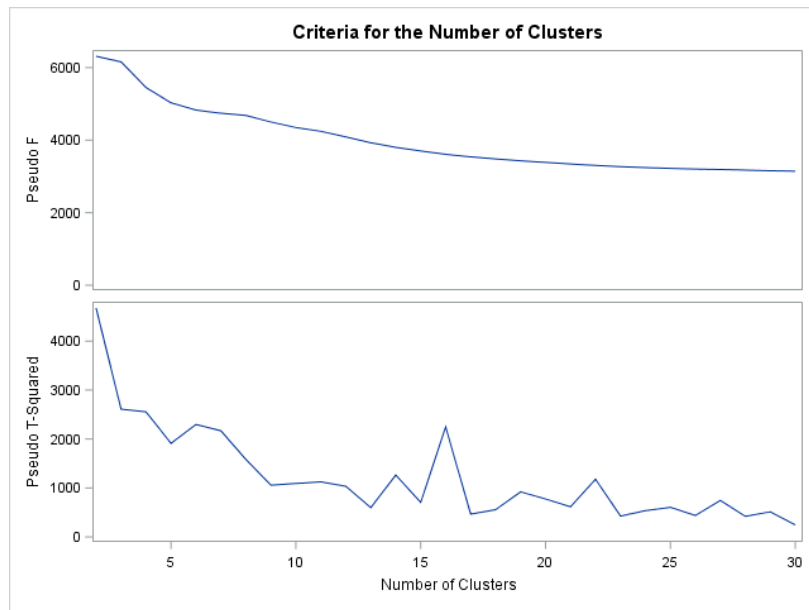


- Prueba 5: 5 variables

Resultado: con 9 grupos explicamos el 64,3%.

SEXO	SIT_PRO	JORNADA	TR_CNAE	BCOTC14
------	---------	---------	---------	---------

Cluster History								
Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square	Pseudo F Statistic	Pseudo t-Squared	Tie
30	CL60	CL36	623	0.0045	.820	3142	246	
29	CL47	CL69	605	0.0045	.816	3158	513	
28	CL33	CL97	1393	0.0047	.811	3174	421	
27	CL34	CL89	1172	0.0051	.806	3190	746	
26	CL63	CL79	523	0.0055	.800	3205	438	
25	CL85	CL27	1629	0.0057	.795	3223	605	
24	CL28	CL67	2125	0.0061	.789	3242	536	
23	CL66	CL45	1070	0.0061	.783	3268	427	
22	CL50	CL107	1147	0.0062	.776	3303	1178	
21	CL105	CL25	2204	0.0065	.770	3342	616	
20	CL91	CL55	744	0.0068	.763	3386	775	
19	CL22	CL88	1897	0.0075	.756	3431	920	
18	CL29	CL21	2809	0.0078	.748	3483	557	
17	CL30	CL43	1368	0.0086	.739	3538	466	
16	CL93	CL115	1432	0.0087	.730	3610	2249	
15	CL20	CL62	1465	0.0087	.722	3702	704	
14	CL90	CL16	2287	0.0097	.712	3802	1261	
13	CL38	CL26	1279	0.0098	.702	3928	598	
12	CL19	CL39	3150	0.0099	.692	4089	1034	
11	CL14	CL35	2981	0.0126	.680	4243	1124	
10	CL23	CL32	2513	0.0179	.662	4347	1092	
9	CL13	CL24	3404	0.0189	.643	4500	1057	
8	CL15	CL58	2365	0.0219	.621	4680	1583	
7	CL18	CL31	4219	0.0338	.587	4740	2168	
6	CL12	CL10	5663	0.0402	.547	4828	2298	
5	CL9	CL17	4772	0.0455	.501	5028	1911	
4	CL11	CL7	7200	0.0516	.450	5450	2555	
3	CL4	CL8	9565	0.0687	.381	6158	2610	
2	CL6	CL3	15228	0.1413	.240	6311	4678	
1	CL2	CL5	20000	0.2399	.000	.	6311	



5. Elección del método de cálculo de distancias, individuos trabajando del clúster laboral.

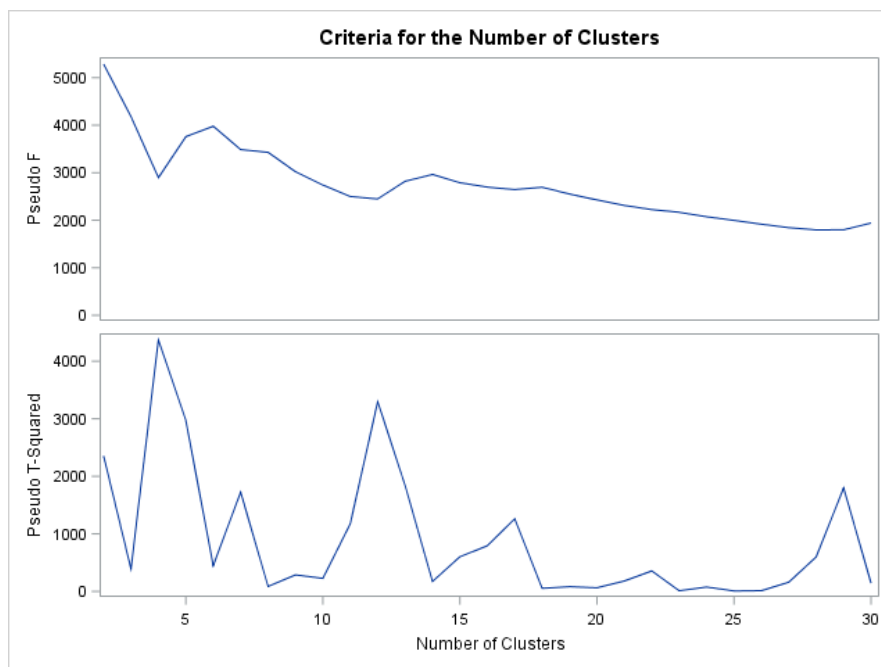
El procedimiento es el mismo que para el clúster académico. Buscamos explicar un 60% de los datos.

- Método de enlace medio

Resultado: 13 grupos explican el 62.9%

Cluster History									
Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square	Pseudo F Statistic	Pseudo t-Squared	Norm RMS Distance	Tie
30	CL33	CL92	831	0.0020	.738	1942	144	0.6727	
29	CL38	CL72	4016	0.0219	.716	1802	1798	0.677	
28	CL39	CL94	2708	0.0081	.708	1796	601	0.6874	
27	CL46	CL177	1444	0.0022	.706	1845	160	0.6874	
26	CL47	CL103	50	0.0001	.706	1918	13.5	0.6897	
25	CL36	CL171	27	0.0001	.706	1997	7.4	0.6939	
24	CL31	CL40	135	0.0009	.705	2075	77.6	0.6958	
23	CL132	CL41	78	0.0002	.705	2168	12.1	0.6988	
22	CL56	CL107	1479	0.0044	.700	2224	356	0.6991	
21	CL54	CL23	1546	0.0020	.698	2313	181	0.725	
20	CL81	CL45	79	0.0007	.698	2427	63.4	0.7302	
19	CL100	CL25	332	0.0008	.697	2552	82.0	0.7526	
18	CL26	CL51	82	0.0007	.696	2693	54.4	0.7647	
17	CL43	CL55	2749	0.0168	.679	2646	1262	0.7752	
16	CL21	CL19	1878	0.0099	.669	2699	794	0.7808	
15	CL64	CL57	1079	0.0080	.661	2790	603	0.7911	

14	CL29	CL32	4100	0.0030	.658	2964	174	0.7935	
13	CL28	CL27	4152	0.0297	.629	2821		0.7964	
12	CL14	CL34	7155	0.0547	.574	2450	3298	0.799	
11	CL30	CL22	2310	0.0185	.556	2500	1176	0.8136	
10	CL35	CL44	381	0.0033	.552	2741	226	0.8325	
9	CL16	CL24	2013	0.0051	.547	3021	287	0.8693	
8	CL20	CL18	161	0.0017	.546	3430	86.5	0.9	
7	CL15	CL17	3828	0.0342	.511	3488	1727	0.9179	
6	CL7	CL10	4209	0.0126	.499	3980	447	0.9385	
5	CL13	CL11	6462	0.0696	.429	3760	2973	0.9683	
4	CL12	CL5	13617	0.1263	.303	2897	4365	0.9776	
3	CL9	CL8	2174	0.0082	.295	4178	394	1.0265	
2	CL4	CL3	15791	0.0856	.209	5286	2355	1.0419	
1	CL2	CL6	20000	0.2091	.000	.	5286	1.1657	

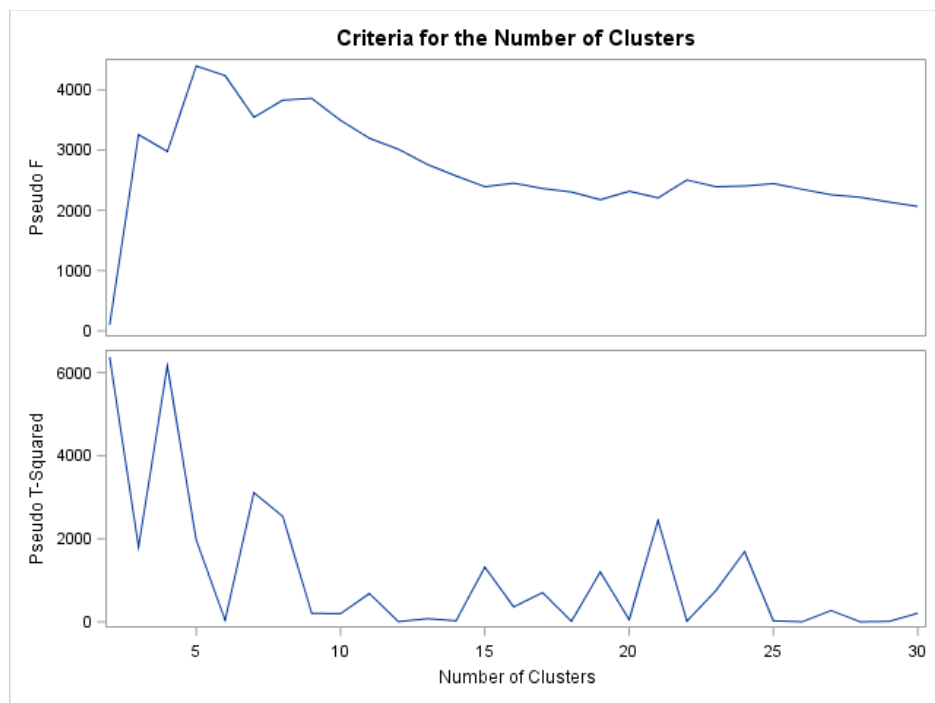


- Método del centroide

Resultado: 12 grupos explican el 62.4%

Cluster History									
Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square	Pseudo F Statistic	Pseudo t-Squared	Norm Centroid Distance	Tie
30	CL121	CL98	925	0.0028	.750	2065	209	0.4924	
29	CL39	CL372	917	0.0001	.750	2138	10.4	0.4958	
28	CL106	CL8274	17	0.0000	.750	2217	2.4	0.4995	
27	CL50	CL84	3655	0.0037	.746	2258	272	0.512	
26	CL37	CL2927	150	0.0000	.746	2349	3.5	0.5161	
25	CL410	CL69	58	0.0001	.746	2445	21.4	0.527	

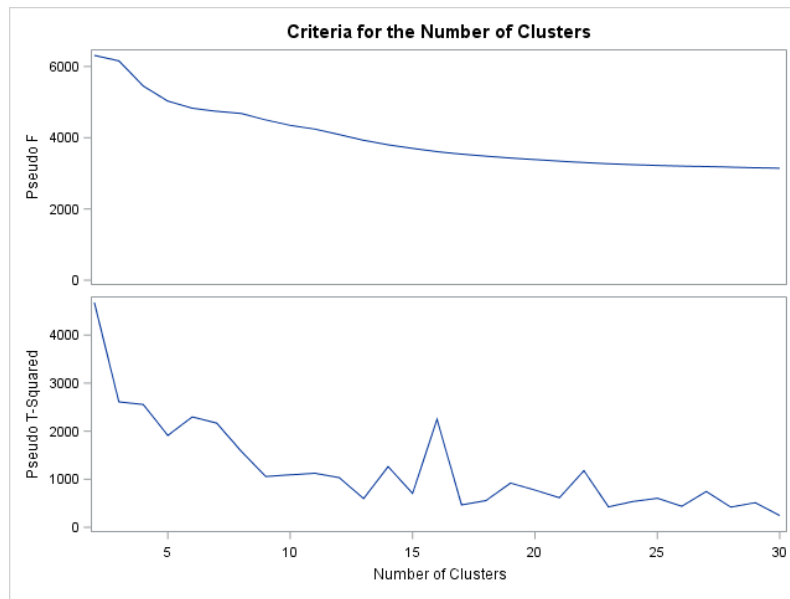
24	CL29	CL49	1628	0.0114	.735	2404	1691	0.5331	
23	CL70	CL32	1818	0.0098	.725	2392	748	0.5463	
22	CL31	CL56	43	0.0002	.725	2504	17.4	0.5514	
21	CL45	CL23	4979	0.0363	.688	2207	2439	0.5605	
20	CL74	CL22	79	0.0006	.688	2316	54.6	0.5644	
19	CL21	CL30	5904	0.0255	.662	2177	1201	0.5715	
18	CL26	CL108	153	0.0001	.662	2304	12.9	0.5822	
17	CL40	CL83	1050	0.0080	.654	2363	703	0.5948	
16	CL17	CL117	1317	0.0064	.648	2450	361	0.5503	
15	CL16	CL92	3342	0.0215	.626	2392	1317	0.5191	
14	CL15	CL28	3359	0.0005	.626	2571	21.7	0.5395	
13	CL14	CL25	3417	0.0018	.624	2764	77.9	0.5561	
12	CL520	CL68	34	0.0001	.624	3015	6.7	0.6073	
11	CL77	CL63	1070	0.0087	.615	3197	682	0.6162	
10	CL11	CL124	1214	0.0039	.611	3494	198	0.5576	
9	CL10	CL18	1367	0.0044	.607	3859	204	0.5682	
8	CL57	CL85	3916	0.0342	.573	3829	2538	0.6213	
7	CL27	CL8	7571	0.0573	.516	3545	3108	0.5504	
6	CL7	CL12	7605	0.0011	.514	4237	41.2	0.5621	
5	CL13	CL9	4784	0.0465	.468	4396	1971	0.6901	
4	CL19	CL6	13509	0.1592	.309	2977	6163	0.692	
3	CL4	CL24	15137	0.0630	.246	3257	1799	0.6586	
2	CL3	CL5	19921	0.2409	.005	96.2	6375	0.8141	
1	CL2	CL20	20000	0.0048	.000	.	96.2	0.7802	



- Método de Ward

Resultado: 9 grupos explican el 64.3%

Cluster History								
Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square	Pseudo F Statistic	Pseudo t-Squared	Tie
30	CL60	CL36	623	0.0045	.820	3142	246	
29	CL47	CL69	605	0.0045	.816	3158	513	
28	CL33	CL97	1393	0.0047	.811	3174	421	
27	CL34	CL89	1172	0.0051	.806	3190	746	
26	CL63	CL79	523	0.0055	.800	3205	438	
25	CL85	CL27	1629	0.0057	.795	3223	605	
24	CL28	CL67	2125	0.0061	.789	3242	536	
23	CL66	CL45	1070	0.0061	.783	3268	427	
22	CL50	CL107	1147	0.0062	.776	3303	1178	
21	CL105	CL25	2204	0.0065	.770	3342	616	
20	CL91	CL55	744	0.0068	.763	3386	775	
19	CL22	CL88	1897	0.0075	.756	3431	920	
18	CL29	CL21	2809	0.0078	.748	3483	557	
17	CL30	CL43	1368	0.0086	.739	3538	466	
16	CL93	CL115	1432	0.0087	.730	3610	2249	
15	CL20	CL62	1465	0.0087	.722	3702	704	
14	CL90	CL16	2287	0.0097	.712	3802	1261	
13	CL38	CL26	1279	0.0098	.702	3928	598	
12	CL19	CL39	3150	0.0099	.692	4089	1034	
11	CL14	CL35	2981	0.0126	.680	4243	1124	
10	CL23	CL32	2513	0.0179	.662	4347	1092	
9	CL13	CL24	3404	0.0189	.643	4500	1057	
8	CL15	CL58	2365	0.0219	.621	4680	1583	
7	CL18	CL31	4219	0.0338	.587	4740	2168	
6	CL12	CL10	5663	0.0402	.547	4828	2298	
5	CL9	CL17	4772	0.0455	.501	5028	1911	
4	CL11	CL7	7200	0.0516	.450	5450	2555	
3	CL4	CL8	9565	0.0687	.381	6158	2610	
2	CL6	CL3	15228	0.1413	.240	6311	4678	
1	CL2	CL5	20000	0.2399	.000	.	6311	



6. Tablas de contribuciones a la inercia y coordenadas, variables y clúster del clúster laboral. Análisis de Correspondencias Simples.

6.1. Variable TR_CNAE y clúster:

- *TR_CNAE*:

Partial Contributions to Inertia for the Row Points				
	Dim1	Dim2	Dim3	Dim4
Actividades administrativas de oficina y otras actividades auxiliares a las empresas	0.0024	0.0007	0.0001	0.0036
Actividades artísticas, culturales, recreativas, deportivas y de entrenamiento	0.0233	0.0001	0.0000	0.0174
Actividades asociativas y sindicales	0.0014	0.0030	0.0023	0.0000
Actividades de agencias viajes, operadores turísticos, servicios reservas y actividades relacionadas	0.0005	0.0003	0.0007	0.0003
Actividades de fotografía	0.0004	0.0004	0.0015	0.0008
Actividades de mantenimiento físico	0.0030	0.0004	0.0000	0.0005
Actividades de organizaciones y organismos extraterritoriales	0.0010	0.0001	0.0000	0.0000
Actividades de seguridad e investigación	0.0005	0.0073	0.0000	0.0008
Actividades financieras y de seguros	0.0256	0.0002	0.0006	0.0000
Actividades inmobiliarias	0.0001	0.0046	0.0021	0.0023
Actividades profesionales, científicas y técnicas	0.0102	0.0174	0.0006	0.0743
Actividades relacionadas con el empleo	0.0055	0.0008	0.0050	0.0001

Actividades sanitarias y de servicios sociales	0.0320	0.1310	0.4447	0.0371
Actividades veterinarias	0.0030	0.0007	0.0012	0.0033
Administración Pública y defensa;Seguridad Social obligatoria	0.0121	0.0079	0.0297	0.6021
Agricultura, ganadería, silvicultura y pesca	0.0016	0.0218	0.0002	0.0033
Comercio al por mayor y al por menor; reparación de vehículos de motor y motocicletas	0.0001	0.0087	0.0385	0.0028
Construcción e ingeniería civil	0.0005	0.0507	0.0001	0.0182
Educación	0.3063	0.0699	0.1984	0.0372
Efectos personales y artículos de uso doméstico y otros servicios personales	0.0014	0.0004	0.0000	0.0005
Hostelería	0.0208	0.0025	0.0006	0.0098
Industria manufacturera	0.0083	0.6439	0.0316	0.0206
Industrias extractivas	0.0009	0.0017	0.0019	0.0011
Información y comunicaciones	0.4989	0.0000	0.2303	0.1432
NS/NC	0.0245	0.0190	0.0081	0.0063
Particulares como empleadores de personal doméstico;particulares:productores de bbss de uso propio	0.0007	0.0005	0.0002	0.0011
Servicios a edificios y actividades de jardinería	0.0000	0.0000	0.0002	0.0050
Suministro de agua, actividades de saneamiento, gestión de residuos y descontaminación	0.0037	0.0005	0.0001	0.0042
Suministro de energía eléctrica, gas, vapor y aire acondicionado	0.0019	0.0057	0.0013	0.0010
Transporte, almacenamiento, actividades postales y de correos	0.0092	0.0001	0.0001	0.0027

- Clúster laboral:

Partial Contributions to Inertia for the Column Points				
	Dim1	Dim2	Dim3	Dim4
7	0.1679	0.0233	0.1118	0.0029
8	0.0019	0.7871	0.0268	0.0044
9	0.0912	0.0794	0.4675	0.0009
10	0.6052	0.0014	0.2538	0.0314
11	0.1016	0.0549	0.0199	0.0207
12	0.0089	0.0022	0.0011	0.0065
13	0.0176	0.0214	0.0000	0.0049
Column Coordinates				
	Dim1	Dim2	Dim3	Dim4
7	-0.3433	0.1210	-0.2310	-0.0244
8	-0.0519	-0.9966	0.1604	-0.0428

9	0.3446	0.3042	0.6436	-0.0182
10	1.2038	0.0539	-0.6429	-0.1496
11	-0.5214	0.3624	-0.1903	-0.1284
12	-0.2671	0.1249	-0.0764	-0.1244
13	-0.3124	0.3254	0.0099	-0.0902
14	0.1420	0.0803	-0.0821	1.0839
15	0.1089	0.5156	0.9070	-0.2754

6.2. SIT_PRO y clúster

- SIT_PRO

Partial Contributions to Inertia for the Row Points			
	Dim1	Dim2	Dim3
Asalariado con contrato permanente	0.5315	0.0564	0.0018
Asalariado con contrato temporal	0.4404	0.1690	0.0420
Ayuda familiar	0.0011	0.0184	0.0091
Empresario o trabajador independiente	0.0025	0.6288	0.2332
En prácticas, formación o becario	0.0245	0.1273	0.7139

Row Coordinates			
	Dim1	Dim2	Dim3
Asalariado con contrato permanente	-0.8381	-0.2094	-0.0306
Asalariado con contrato temporal	0.8296	-0.3942	-0.1605
Ayuda familiar	0.2702	0.8379	-0.4802
Empresario o trabajador independiente	0.1153	1.4141	-0.7030
En prácticas, formación o becario	0.3143	0.5498	1.0628

- Clúster laboral (trabajando):

Partial Contributions to Inertia for the Column Points			
	Dim1	Dim2	Dim3
7	0.1490	0.0028	0.0064
8	0.0097	0.6481	0.1383
9	0.1172	0.0744	0.0065
10	0.0106	0.0682	0.8141

11	0.0191	0.0562	0.0033
12	0.0919	0.0600	0.0224
13	0.1451	0.0215	0.0002
14	0.2180	0.0257	0.0063
15	0.2393	0.0431	0.0024

Column Coordinates			
	Dim1	Dim2	Dim3
7	0.5231	0.0553	-0.0681
8	-0.2137	1.3379	-0.5046
9	0.7380	-0.4509	-0.1087
10	0.2998	0.5831	1.6440
11	0.3629	-0.4773	-0.0949
12	1.1268	-0.6981	-0.3481
13	-1.0782	-0.3180	0.0223
14	-0.9900	-0.2609	0.1056
15	-1.1241	-0.3659	-0.0711

7. Selección de variables del clúster laboral los desempleados que si han trabajado alguna vez.

De nuevo, el procedimiento de selección de variables es el mismo. En este caso, la muestra no está estratificada,

Las variables con las que contamos son las siguientes:

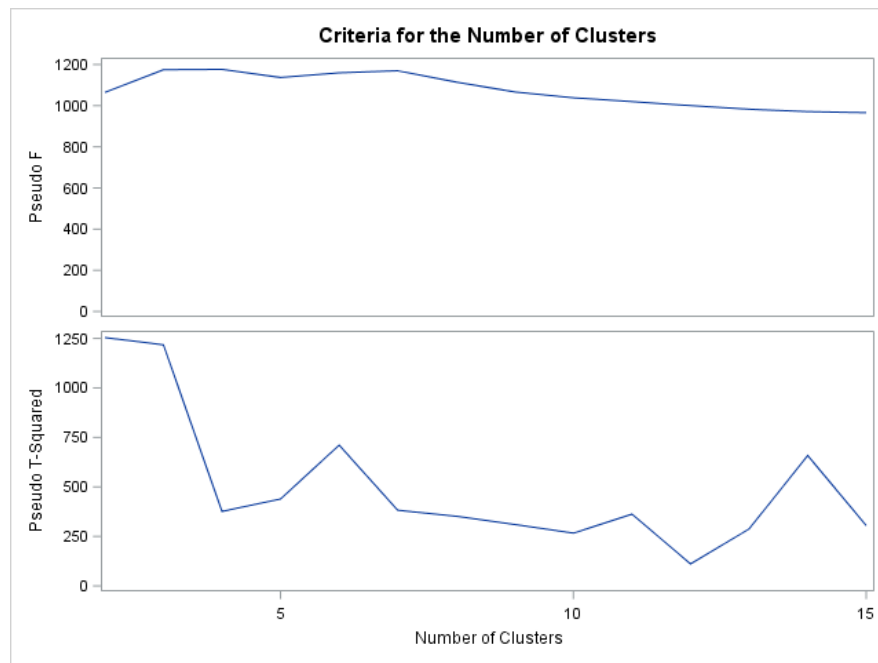
Variable	Descripción
BQ_NUM	Número de empleadores distintos que ha tenido
BQ_TP_T	Tiempo que ha estado trabajando desde que terminó los estudios
TRAB_F1	Ha tenido algún trabajo remunerado durante la realización de los estudios
BSQRE	Ha rechazado algún trabajo por considerarlo poco adecuado desde que se ha titulado
SEXO	

- Prueba 1: 5 variables, todas las anteriores.

Resultado: 7 grupos consiguen explicar el 62,9%.

Cluster History							
Number of Clusters	Clusters Joined	Freq	Semipartial R-Square	R-Square	Pseudo F Statistic	Pseudo t-Squared	Tie

15	CL16	CL53	471	0.0107	.766	966	305	
14	CL48	CL44	426	0.0126	.753	971	659	
13	CL15	CL54	703	0.0129	.740	983	288	
12	CL19	CL22	266	0.0137	.727	1000	111	
11	CL14	CL35	650	0.0154	.711	1020	363	
10	CL42	CL28	413	0.0183	.693	1038	266	
9	CL11	CL20	951	0.0197	.673	1067	310	
8	CL13	CL21	989	0.0201	.653	1115	352	
7	CL26	CL17	608	0.0245	.629	1169	382	
6	CL18	CL23	924	0.0455	.583	1160	711	
5	CL12	CL10	679	0.0601	.523	1137	439	
4	CL5	CL7	1287	0.0633	.460	1177	376	
3	CL9	CL8	1940	0.0982	.362	1175	1218	
2	CL3	CL6	2864	0.1574	.204	1065	1254	
1	CL4	CL2	4151	0.2042	.000	.	1065	

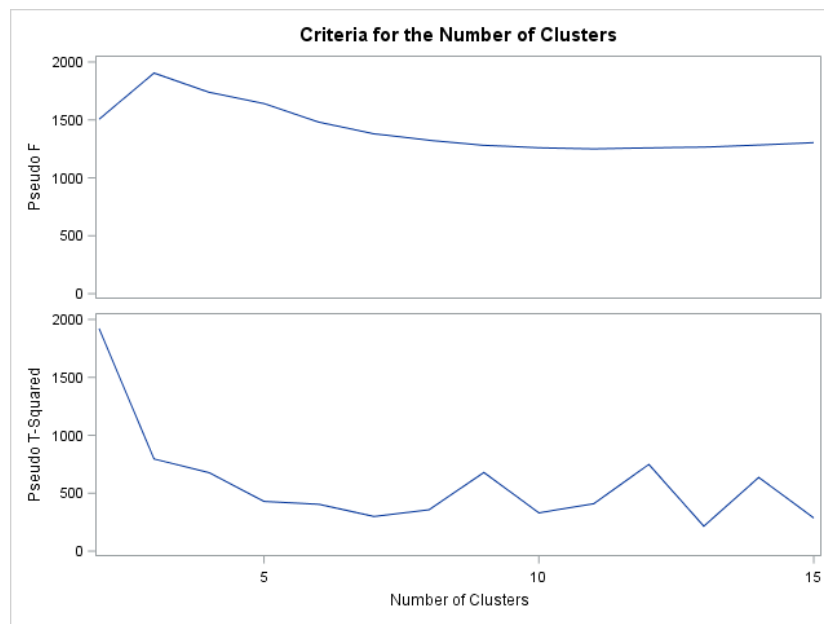


- Prueba 2: 4 variables

Resultado: 5 grupos consiguen explicar el 61,3%

SEXO		BQ_TP_T		BQ_NUM		BSQRE		
Cluster History								
Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square	Pseudo F Statistic	Pseudo t-Squared	Tie
15	CL22	CL19	440	0.0138	.815	1304	286	
14	CL33	CL30	439	0.0140	.801	1283	636	
13	CL45	CL18	564	0.0154	.786	1265	215	
12	CL32	CL38	403	0.0160	.770	1258	749	

11	CL46	CL14	629	0.0186	.751	1250	410	
10	CL21	CL16	826	0.0188	.732	1259	330	
9	CL40	CL27	471	0.0202	.712	1281	680	
8	CL29	CL15	697	0.0210	.691	1325	357	
7	CL8	CL11	1326	0.0247	.666	1380	300	
6	CL23	CL9	729	0.0255	.641	1480	404	
5	CL10	CL28	1129	0.0280	.613	1642	429	
4	CL12	CL5	1532	0.0559	.557	1739	678	
3	CL13	CL6	1293	0.0783	.479	1905	796	
2	CL7	CL4	2858	0.2125	.266	1506	1922	
1	CL2	CL3	4151	0.2663	.000	.	1506	



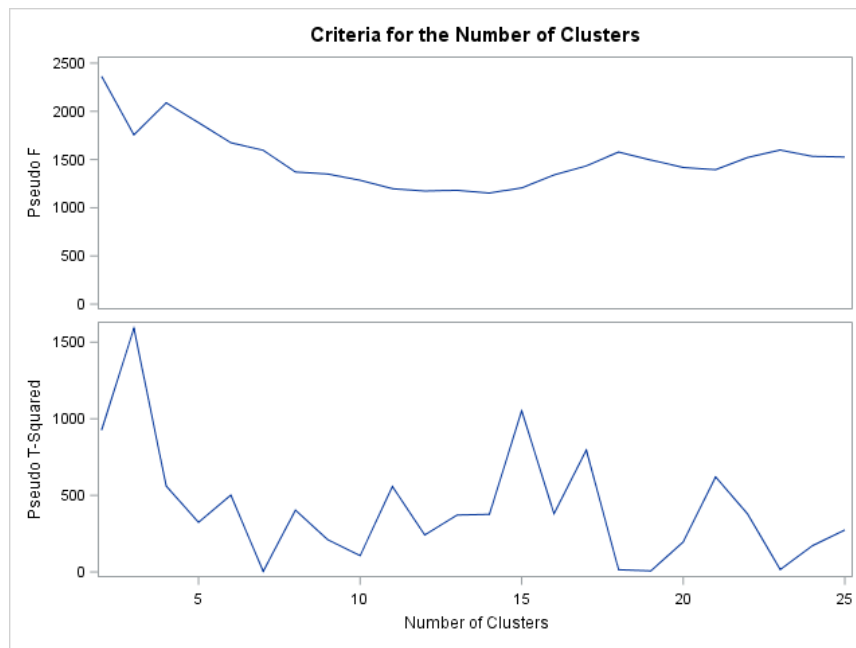
8. Elección del método de cálculo de distancias de desempleados que si han trabajado dentro del clúster laboral.

- Método de enlace medio

Resultado: 4 grupos explican el 60.2%.

Cluster History									
Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square	Pseudo F Statistic	Pseudo t-Squared	Norm RMS Distance	Tie
25	CL92	CL27	202	0.0037	.899	1526	275	0.4851	
24	CL90	CL29	239	0.0036	.895	1533	172	0.5171	
23	CL69	CL2597	181	0.0002	.895	1598	15.8	0.5393	
22	CL25	CL23	383	0.0093	.886	1523	376	0.5503	

21	CL35	CL37	444	0.0146	.871	1394	620	0.6116	
20	CL39	CL32	101	0.0040	.867	1418	196	0.6523	
19	CL38	O8604	141	0.0002	.867	1495	7.2	0.6533	
18	CL58	CL28	575	0.0003	.867	1578	14.2	0.6668	
17	CL43	CL31	470	0.0192	.847	1434	796	0.6669	
16	CL21	CL36	637	0.0179	.829	1340	380	0.6677	
15	CL18	CL49	791	0.0262	.803	1206	1052	0.6705	
14	CL62	CL17	697	0.0196	.784	1152	376	0.6739	
13	CL33	CL19	237	0.0096	.774	1180	371	0.6756	
12	CL14	CL55	877	0.0166	.757	1174	241	0.677	
11	CL47	CL42	348	0.0142	.743	1198	558	0.6791	
10	CL13	CL20	338	0.0069	.736	1284	106	0.6795	
9	CL30	CL10	522	0.0135	.723	1350	210	0.6807	
8	CL15	CL11	1139	0.0244	.698	1370	404	0.6833	
7	CL12	CL26	880	0.0004	.698	1596	4.0	0.689	
6	CL56	CL16	988	0.0292	.669	1674	502	0.6931	
5	CL8	CL24	1378	0.0239	.645	1882	324	0.7033	
4	CL6	CL22	1371	0.0431	.602	2088	561	0.7796	
3	CL5	CL7	2258	0.1431	.459	1757	1591	0.9616	
2	CL4	CL9	1893	0.0956	.363	2364	926	0.9671	
1	CL2	CL3	4151	0.3630	.000	.	2364	1.1698	

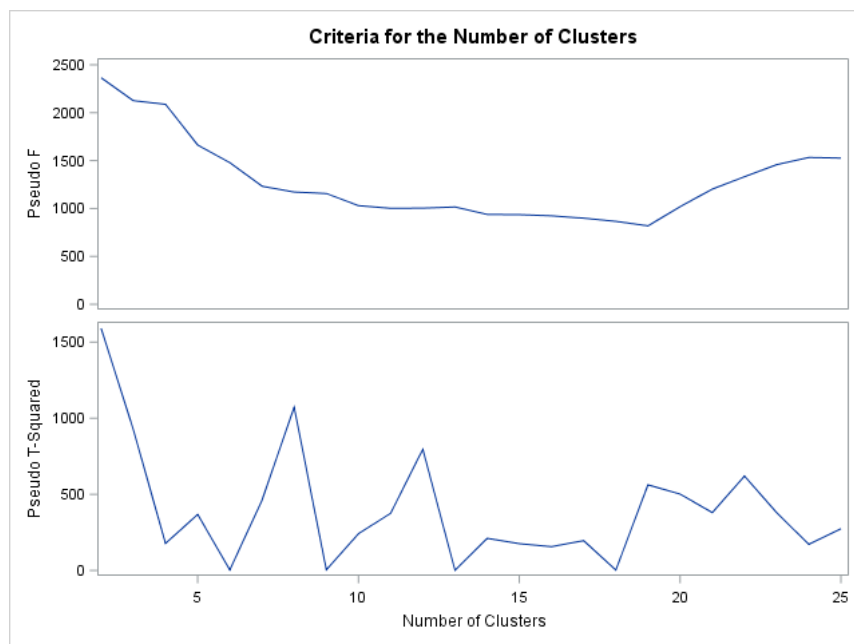


- Método del centroide, y el escogido:

Resultado: 4 grupos explican el 60.2%.

Cluster History

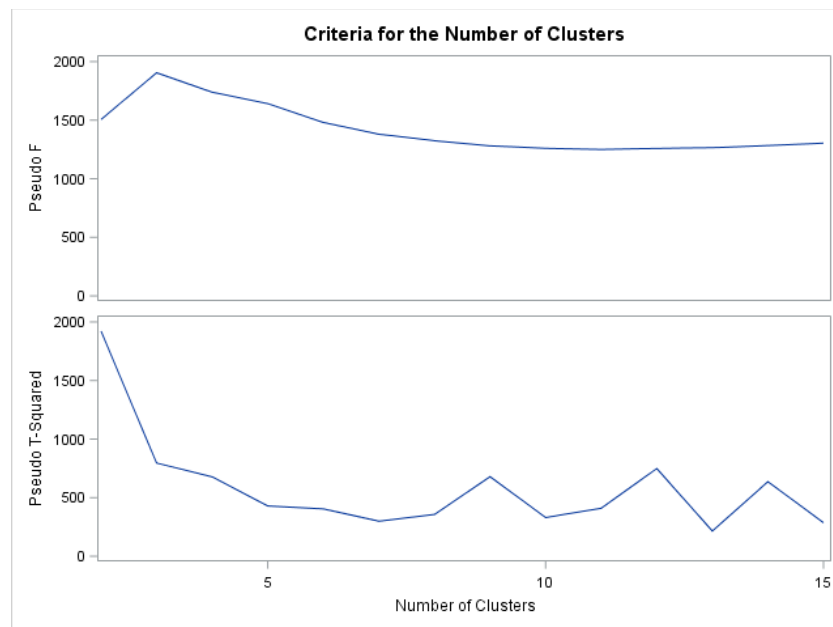
Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square	Pseudo F Statistic	Pseudo t-Squared	Norm Centroid Distance	Tie
25	CL90	CL44	202	0.0037	.899	1526	275	0.4266	
24	CL85	CL72	239	0.0036	.895	1533	172	0.4383	
23	CL25	CL46	381	0.0092	.886	1458	380	0.4487	
22	CL51	CL29	444	0.0146	.871	1332	620	0.5253	
21	CL22	CL34	637	0.0179	.853	1203	380	0.5257	
20	CL68	CL21	988	0.0292	.824	1020	502	0.5176	
19	CL20	CL23	1369	0.0432	.781	819	563	0.5712	
18	CL19	CL2597	1371	0.0003	.781	866	2.4	0.5214	
17	CL41	CL27	101	0.0040	.777	899	196	0.5837	
16	CL75	CL17	197	0.0067	.770	923	156	0.533	
15	CL16	CL61	337	0.0098	.760	937	176	0.4986	
14	CL81	CL15	521	0.0135	.747	938	211	0.4847	
13	CL14	O8604	522	0.0001	.747	1016	1.5	0.5327	
12	CL54	CL47	470	0.0192	.727	1004	796	0.5872	
11	CL86	CL12	697	0.0196	.708	1003	376	0.516	
10	CL11	CL78	877	0.0166	.691	1030	241	0.4912	
9	CL10	CL32	880	0.0004	.691	1157	4.0	0.495	
8	CL91	CL58	789	0.0263	.665	1172	1073	0.5896	
7	CL8	CL65	1003	0.0235	.641	1233	461	0.5381	
6	CL7	CL26	1005	0.0002	.641	1479	3.3	0.5084	
5	CL6	CL24	1244	0.0247	.616	1663	368	0.515	
4	CL5	CL37	1378	0.0144	.602	2088	178	0.4966	
3	CL18	CL13	1893	0.0956	.506	2125	926	0.7243	
2	CL4	CL9	2258	0.1431	.363	2364	1591	0.7437	
1	CL3	CL2	4151	0.3630	.000	.	2364	0.8552	



- Método de Ward

Resultado: 4 grupos explican el 55.7%

Cluster History								
Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square	Pseudo F Statistic	Pseudo t-Squared	Tie
15	CL22	CL19	440	0.0138	.815	1304	286	
14	CL33	CL30	439	0.0140	.801	1283	636	
13	CL45	CL18	564	0.0154	.786	1265	215	
12	CL32	CL38	403	0.0160	.770	1258	749	
11	CL46	CL14	629	0.0186	.751	1250	410	
10	CL21	CL16	826	0.0188	.732	1259	330	
9	CL40	CL27	471	0.0202	.712	1281	680	
8	CL29	CL15	697	0.0210	.691	1325	357	
7	CL8	CL11	1326	0.0247	.666	1380	300	
6	CL23	CL9	729	0.0255	.641	1480	404	
5	CL10	CL28	1129	0.0280	.613	1642	429	
4	CL12	CL5	1532	0.0559	.557	1739	678	
3	CL13	CL6	1293	0.0783	.479	1905	796	
2	CL7	CL4	2858	0.2125	.266	1506	1922	
1	CL2	CL3	4151	0.2663	.000	.	1506	



9. Tablas de contribuciones a la inercia y coordenadas, variables y clúster del clúster laboral de los desempleados. Análisis de Correspondencias Simples.

9.1. BQ_NUM y clúster

- BQ_NUM

	Dim1	Dim2	Dim3
4 o más	0.0334	0.2162	0.4621
Dos	0.0230	0.2433	0.4775
NS/NC	0.0031	0.0008	0.0056
Tres	0.4885	0.2682	0.0361
Uno	0.4520	0.2714	0.0188

Row Coordinates			
	Dim1	Dim2	Dim3
4 o más	0.1835	-0.3713	0.3483
Dos	-0.1601	-0.4138	-0.3720
NS/NC	-0.3275	0.1367	0.2251
Tres	0.8199	0.4831	-0.1136
Uno	-0.7238	0.4460	0.0753

- Clúster:

Partial Contributions to Inertia for the Column Points			
	Dim1	Dim2	Dim3
3	0.0463	0.1884	0.2245
4	0.0460	0.0441	0.6866
5	0.4151	0.4605	0.0254
6	0.4926	0.3070	0.0636

Column Coordinates			
	Dim1	Dim2	Dim3
3	-0.1562	-0.2506	-0.1755
4	0.2423	-0.1887	0.4776
5	-1.0933	0.9155	0.1379
6	1.0131	0.6359	-0.1857

9.2. BQ_TP_P y clúster

Partial Contributions to Inertia for the Row Points			
	Dim1	Dim2	Dim3
2 o más años	0.0376	0.0073	0.5741
De 1 año a año y medio	0.8131	0.0155	0.0021
De 1 año y medio a 2 años	0.0095	0.0057	0.2389
De 6 meses a 1 año	0.0337	0.3724	0.1272
Menos de 6 meses	0.1059	0.5879	0.0537
NS/NC	0.0002	0.0111	0.0040

	Dim1	Dim2	Dim3
2 o más años	-0.1538	-0.0568	-0.3086
De 1 año a año y medio	1.0728	0.1243	0.0278
De 1 año y medio a 2 años	-0.1221	-0.0796	0.3136
De 6 meses a 1 año	-0.1971	-0.5497	0.1964
Menos de 6 meses	-0.3993	0.7894	0.1459
NS/NC	0.1869	1.3479	0.4956

- Clúster:

Partial Contributions to Inertia for the Column Points			
	Dim1	Dim2	Dim3
3	0.0943	0.0008	0.3640
4	0.0110	0.3337	0.4319
5	0.0437	0.6538	0.2035
6	0.8510	0.0116	0.0006

Column Coordinates		
Dim1	Dim2	Dim3
-0.1975	-0.0157	-0.1991
-0.1050	-0.4852	0.3375
-0.3142	1.0199	0.3479
1.1794	0.1157	-0.0157

7. Análisis de correspondencias simples:

7.1. Tabla de contingencia del estudio de las proporciones sobre el total de la población de cada modalidad.

Contingency Table											
Perce nts	ACAD _1	ACAD _10	ACAD _2	ACAD _3	ACAD _4	ACAD _5	ACAD _6	ACAD _7	ACAD _8	ACAD _9	Sum
LAB_ 1	1.249	0.183	1.557	0.751	0.547	0.000	0.204	1.280	0.557	0.339	6.667
LAB_ 10	1.574	0.003	0.477	0.865	0.996	0.000	0.042	0.429	0.093	0.277	4.757
LAB_ 11	2.065	0.045	0.073	1.204	1.512	0.000	0.138	0.093	0.166	0.550	5.847
LAB_ 12	0.858	0.000	0.000	1.024	0.554	0.000	0.031	0.000	0.093	0.356	2.916
LAB_ 13	0.329	0.055	1.830	0.263	0.249	0.000	0.180	1.446	0.533	0.145	5.030
LAB_ 14	0.893	0.135	3.083	0.343	0.702	0.000	0.571	2.097	0.906	0.232	8.960
LAB_ 15	3.636	0.000	0.045	0.934	1.879	0.000	0.086	0.031	0.059	0.962	7.632
LAB_ 2	0.502	0.816	0.000	0.519	0.484	2.086	0.024	0.000	0.163	0.052	4.646
LAB_ 3	0.405	1.439	0.000	0.298	0.464	4.764	0.086	0.000	0.249	0.062	7.767
LAB_ 4	0.581	0.332	0.000	0.626	0.599	0.924	0.021	0.000	0.048	0.076	3.207
LAB_ 5	0.339	0.156	0.000	0.221	0.256	0.360	0.003	0.000	0.059	0.028	1.422
LAB_ 6	0.031	0.436	0.000	0.031	0.086	1.277	0.028	0.000	0.076	0.000	1.965
LAB_ 7	1.007	0.374	7.269	1.142	0.882	0.000	1.159	6.860	2.674	0.567	21.934
LAB_ 8	3.563	0.021	0.142	1.771	1.477	0.000	0.111	0.159	0.221	1.114	8.580

LAB_9	0.723	0.156	2.190	0.737	0.547	0.000	0.287	2.726	0.993	0.311	8.670
Sum	17.755	4.152	16.665	10.728	11.233	9.410	2.972	15.122	6.892	5.072	100.000



ENCUESTA DE INSERCIÓN LABORAL DE LOS TITULADOS UNIVERSITARIOS 2014

CAWI

CUESTIONARIO

LEYENDA:

Textos en NEGRO: Corresponden a las preguntas

Textos con fondo GRIS. Corresponden a Notas y aclaraciones

Notas en ROJO: Se refieren a los flujos que debe seguir el cuestionario según las diferentes respuestas.

Primera Pantalla:

1- Breve Introducción a la encuesta

Muchas gracias por tu colaboración. Has entrado en la plataforma de cumplimentación de la Encuesta de Inserción Laboral de Titulados Universitarios.

Este trabajo es el primer estudio de inserción laboral de los titulados del Sistema Universitario realizado por el Instituto Nacional de Estadística. El objetivo principal es

conocer la situación profesional y los diversos aspectos del proceso de inserción laboral (acceso al mercado de trabajo) de los titulados universitarios.

Te rogamos que cumplimentes el siguiente cuestionario de forma exacta y veraz.

2- Instrucciones de cumplimentación

Lee atentamente cada pregunta. Marca la opción de respuesta adecuada para cada una de ellas y luego pulsa "Siguiente".

En cualquier momento puedes interrumpir la cumplimentación para continuar en otro momento desde la última pregunta que hayas contestado, simplemente pulsando en "Salir".

Existe un teléfono gratuito, el 900 XXX XXX, donde puedes contactar si tienes dudas sobre la cumplimentación del cuestionario u otro tipo de problemas que te impidan la cumplimentación.

INTRODUCCIÓN

Según los datos de los que disponemos, acabaste la carrera de "..." en 2010.

P1: ¿Son correctos estos datos?

- Sí ☐ 1 → ir a P4
- No ☐ 2

P2: ¿Por qué motivo no son correctos?

- No realicé los estudios de "..." ☐ 1 → FIN DE CUESTIONARIO
- Terminé la carrera de "..." pero en otro año ☐ 2

P3: ¿En qué año la terminaste?

- Año _____

Si P3 ≠ (2008, 2009, 2010, 2011, 2012) → FIN DE CUESTIONARIO

1. EDUCACIÓN Y APRENDIZAJE DEL TITULADO UNIVERSITARIO

1.1. Formación completada

Comenzaremos con las siguientes preguntas relacionadas con los estudios que realizaste para obtener el título de "..."

P4: ¿Disfrutaste de alguna beca durante tus estudios para obtener el título de "..."?

- Sí _____ ☐ 1
- No _____ ☐ 2 → ir a P6

P5: ¿Cuál o cuáles de las siguientes becas has disfrutado?

	Si marca esta opción	Si NO marca esta opción
a. Beca general para el estudio	1	2
b. Premio o beca de excelencia	1	2
c. Beca de colaboración en la Universidad	1	2
d. Beca de prácticas externas, para prestar un servicio o desempeñar un trabajo fuera de la Universidad	1	2
e. Beca para realizar estudios fuera de España (Incluidas becas complemento a programas Erasmus)	1 → ir a P7	2

P6: ¿Has realizado parte de tus estudios para obtener el título de "..." fuera de España?

- Sí _____ ☐ 1 → ir a P8 - No _____ ☐ 2 → ir a P10

En el caso de haber participado en varios programas de movilidad fuera de España, refiere las siguientes preguntas al de mayor duración.

P7: Por favor, indica el programa o beca de movilidad que disfrutaste

- Programa Erasmus_____ ☐ 1
- Otros programas o becas dentro de la UE ☐ 2
- Otros programas o becas fuera de la UE ☐ 3

P8: Por favor, señala el país o región de la universidad de destino.

- País o región___ ☐ Desplegable D4.1

P9: ¿Durante cuántos meses?

- Meses_____ ☐ ☐

P10: Además de los estudios realizados en "universidad...", ¿has realizado parte de tus estudios para obtener el título de "..." en otra universidad española?

- Sí_____ ☐ 1
- No_____ ☐ 2

P11: Si tuvieras que empezar de nuevo, ¿volverías a cursar estudios universitarios?

- Sí_____ ☐ 1
- No_____ ☐ 2 → ir a P13

P12: ¿Volverías a estudiar la misma titulación? (Recuerda que todas las preguntas están referidas a la misma titulación)

- Sí_____ ☐ 1
- No_____ ☐ 2

1.2. Otra formación realizada

A continuación, necesitamos que respondas a algunas preguntas sobre otra formación distinta de la que realizaste para obtener el título de "..."

P13: Aparte de los estudios para obtener el título de "...", indica si has completado alguno de los siguientes estudios. (Marca tantas opciones como sea necesario)

	Si marca esta opción	Si NO marca esta opción
a. Otros estudios Universitarios	1	2
b. Estudios artísticos superiores (artes plásticas, diseño, música y danza, arte dramático o conservación y restauración de bienes culturales)	1 → ir a P17	2
c. Ciclos Formativos de grado superior (formación profesional, artes plásticas y diseño) o Enseñanzas Deportivas de Grado Superior	1 → ir a P17	2
d. No he completado ninguno de los estudios anteriores	1 → ir a P17	2

P14: ¿Has cursado parte de estos estudios universitarios fuera de España? (Sólo si P13a = 1)

- Sí _____ ☐ 1
- No _____ ☐ 2 → ir a P17

P15: ¿Qué tipo de estudios universitarios, de los que aparecen a continuación, has completado? (Si has cursado más de un estudio universitario fuera de España, haz referencia al más reciente)

- Diplomatura, ingeniería técnica, arquitectura técnica _____ ☐ 1
- Licenciatura, ingeniería, arquitectura. _____ ☐ 2
- Grado universitario de 240 créditos _____ ☐ 3
- Grado universitario de más de 240 créditos _____ ☐ 4
- Máster _____ ☐ 5
- Doctorado universitario _____ ☐ 6

P16: ¿Cuándo terminaste estos estudios? (Si has cursado más de un estudio universitario fuera de España, haz referencia al más reciente)

- Antes de comenzar los estudios para obtener el título de "..."☐ 1
- Durante la realización de los estudios para obtener el título de "..."☐ 2
- Después de obtener el título de "..."☐ 3

P17 ¿Has completado un máster en una universidad española?

- Sí _____ ☐ 1
- No _____ ☐ 2

Si P17 = 2 y tiene una titulación que permite acceder a formación sanitaria especializada → ir a P21

Si P17 = 2 y no tiene una titulación que permite acceder a formación sanitaria especializada → ir a P22

P18: Por favor, señala la Universidad donde cursaste estos estudios de Máster. (Si has realizado más de un máster, haz referencia al más reciente)

- Universidad___ ☐ Desplegable D6

P19: Por favor, señala la rama de conocimiento de estos estudios de Máster. (Si has realizado más de un máster, haz referencia al más reciente)

- Ciencias_____ ☐ 1
- Ciencias de la Salud___ ☐ 2
- Ciencias sociales y jurídicas _____ ☐ 3
- Arte y Humanidades___ ☐ 4
- Ingeniería y Arquitectura____ ☐ 5

P20: ¿Cuándo has terminado este máster?

- Antes de comenzar los estudios para obtener el título de "..." _____ ☐ 1
- Durante la realización de los estudios para obtener el título de "..."_ ☐ 2
- Después de obtener el título de "..." _____ ☐ 3

Si no tiene una titulación que permite acceder a formación sanitaria especializada → ir a P22

P21: ¿Has realizado o estás realizando alguna formación sanitaria especializada (MIR, EIR, FIR, PIR, BIR, RFIR, QIR, etc.)? (Sólo se contesta si está en el SIU con una titulación de tipo sanitario, tanto si es la titulación del curso de referencia como una titulación anterior o posterior).

- Sí, ya he completado esta formación. _____ ☐ 1
- Sí, estoy realizando el MIR o EIR o FIR o PIR, etc. _____ ☐ 2
- No, pero estoy preparando el examen de acceso _____ ☐ 3
- No _____ ☐ 4

P22 Además de tus estudios universitarios, ¿completaste algún estudio de grado medio de formación profesional, artes plásticas y diseño o enseñanzas deportivas?

- Sí _____ ☐ 1
- No _____ ☐ 2

P23: Si estás realizando ahora alguno de los siguientes estudios, por favor, indica cuál (En caso de que estés realizando más de uno, señala el estudio al que dediques más tiempo)

- Grado universitario de 240 créditos _____ ☐ 1
- Grado universitario de más de 240 créditos _____ ☐ 2
- Máster _____ ☐ 3
- Doctorado universitario _____ ☐ 4
- Estudios artísticos superiores (artes plásticas, diseño, música y danza, arte dramático o conservación y restauración de bienes culturales) _____ ☐ 5
- Ciclos Formativos de grado superior (formación profesional, artes plásticas y diseño) o Enseñanzas Deportivas de Grado Superior _____ ☐ 6
- No estoy cursando ninguno de los estudios anteriores _____ ☐ 7

P24: ¿Estás realizando ahora algún tipo de estudios o formación fuera de los planes oficiales de estudio? (cursos impartidos por academias, colegios profesionales, sindicatos, cursos en el centro de trabajo, cursos dirigidos a los desempleados, seminarios, conferencias, clases particulares recibidas, etc.)

- Sí _____ ☐ 1
- No _____ ☐ 2

Para completar la información que nos has facilitado te preguntaremos ahora sobre los idiomas que conoces y puedes usar, y sobre tus habilidades en el uso del ordenador u otros dispositivos informáticos. Empecemos por tu/s lengua/s materna/s, es decir, la/s lengua/s que usabas en tu infancia.

P25: ¿Cuál o cuáles son tus lenguas maternas?

IDIOMA 1 _____ ☐ Desplegable D1

IDIOMA 2 _____ ☐ Desplegable D1

P26: Sin tener en cuenta tu/s lengua/s materna/s ¿cuántos idiomas distintos de tu/s lengua/s materna/s puedes usar?

- Número de idiomas _____ ☐ 1-7
- Ninguno _____ ☐ 0 → Ir a P28

P27 Por favor, indica cuáles son estos idiomas y el nivel de conocimiento de los mismos

IDIOMAS	Nivel alto (Entiendo una gran variedad de textos complicados y uso el idioma con flexibilidad)	Nivel medio (Entiendo lo esencial en el lenguaje común y redacto textos sencillos)	Nivel bajo (Entiendo y uso las expresiones más comunes)
Idioma 1: Desplegable D1	1	2	3
Idioma 2: Desplegable D1	1	2	3
Idioma 3: Desplegable D1	1	2	3
Idioma 4: Desplegable D1	1	2	3
Idioma 5: Desplegable D1	1	2	3
Idioma 6: Desplegable D1	1	2	3
Idioma 7: Desplegable D1	1	2	3

P28: Si tienes algún título oficial de idiomas, por favor, indica cuál/es. (Marca tantas opciones como sea necesario)

	Si marca esta opción	Si NO marca esta opción
Título/s de la Escuela Oficial de Idiomas	1	2
Título/s en lenguas extranjeras reconocidos oficialmente (PET, FCE, CAE, CPE, DELF, DALF, etc.)	1	2
Título/s en lenguas nacionales reconocidos oficialmente de cualquier nivel (A1, A2, B1, B2, C1, C2)	1	2
No tengo ningún título oficial de idiomas	1	2

P29: Por favor, elige la alternativa que mejor describe tu capacidad para usar el ordenador u otros dispositivos informáticos

- Usuario de nivel básico (navegar por internet, enviar correos electrónicos, copiar o mover archivos o carpeta, escribir un texto usando un procesador de textos, usar fórmulas simples en hojas de cálculo) _____ ☐ 1
- Usuario de nivel avanzado (dar formato a textos, usar fórmulas más avanzadas y crear gráficos en hojas de cálculo, instalar dispositivos y/o programas, usar bases de datos) _____ ☐ 2
- Usuario experto (escribir macros, programar, resolver problemas de software y hardware cuando el ordenador no funciona adecuadamente) _____ ☐ 3
- No sé _____ ☐ 9

1.3. Prácticas laborales realizadas durante su formación

Por favor, responde a continuación a las siguientes preguntas sobre las prácticas en empresas realizadas durante tu formación universitaria.

P30: ¿Realizaste prácticas en empresas, instituciones o similares como parte del plan de estudios de tu carrera? (Si has cursado más de unos estudios ten en cuenta los relacionados con la titulación de "...")

- Sí _____ ☐ 1
- No _____ ☐ 2 → ir a P32

P31: ¿Durante cuántos meses?

- Meses _____ ☐ ☐

P32: ¿Realizaste prácticas en empresas, instituciones o similares que no formasen parte del plan de estudios de tu carrera? (Si has cursado más de unos estudios ten en cuenta los relacionados con la titulación de "...")

- Sí _____ ☐ 1
- No _____ ☐ 2 → ir a P34

P33: ¿Durante cuántos meses?

- Meses _____ ☐ ☐

2. PROCESO DE INCORPORACIÓN AL MERCADO LABORAL

2.1. Trabajos realizados durante su formación

Ahora necesitamos que respondas a las siguientes preguntas sobre tu experiencia laboral durante el periodo de formación. Recuerda que estas preguntas hacen referencia a la titulación que te hemos mencionado al inicio de este cuestionario, es decir, tu título de "...".

P34: Aparte de si has realizado prácticas en empresa, ¿has tenido algún trabajo remunerado mientras estudiabas la carrera (incluido el trabajo por cuenta propia, trabajos de formación, en prácticas, puestos de becario, etc.)?

- Sí _____ ☐ 1
- No _____ ☐ 2 → ir a P41

P35: ¿Alguno de esos trabajos estaba relacionado con tu formación para obtener el título de "..."?

- Sí _____ ☐ 1
- No _____ ☐ 2
- No sé__ ☐ 9

En el caso de que hayas tenido varios trabajos, responde la siguiente pregunta pensando en el último que hayas tenido durante tus estudios para obtener el título de "...".

P36: ¿Continuaste en este trabajo durante al menos 6 meses después de acabar la carrera?

- Sí _____ ☐ 1
- No _____ ☐ 2

P37: ¿Cómo era este trabajo?

- Trabajo esporádico (durante menos de tres meses) _____ ☐ 1
- Trabajo a jornada Parcial (durante más de tres meses) _____ ☐ 2
- Trabajo a jornada Completa (durante más de tres meses) _ ☐ 3

Si P36 = 1 → ir a P38 Si P36 = 2 → ir a P40

2.2. Empleo posterior a su titulación

Las siguientes preguntas son sobre la transición al mundo laboral. Recuerda que estas preguntas hacen referencia a la titulación que te hemos mencionado al inicio de este cuestionario, es decir, tu título de "...".

P38: Una vez completados estos estudios, ¿mejoraron las condiciones del trabajo en el que continuaste después de acabar la carrera?

- Sí _____ ☐ 1
- No _____ ☐ 2

P39: Mientras tenías este empleo, ¿buscaste un trabajo mejor?

- Sí _____ ☐ 1

- No _____ ☐ 2

Ir a P45

P40: ¿Has tenido algún otro trabajo remunerado desde que acabaste la carrera?

- Sí _____ ☐ 1 → ir a P42 - No _____ ☐ 2 → ir a P46

P41: ¿Has tenido algún trabajo remunerado desde que acabaste la carrera? (No olvides incluir el trabajo por cuenta propia, trabajos de formación, en prácticas, puestos de becario, etc.)

- Sí _____ ☐ 1 → ir a P42 - No _____ ☐ 2 → ir a P46

2.3. Procesos de búsqueda de empleo posterior a su titulación

Por favor, responde a continuación a algunas preguntas sobre la manera en la que buscaste empleo después de obtener tu título de "...".

P42: ¿Cuánto tiempo pasó desde que acabaste la carrera hasta que empezaste en este trabajo?

- Menos de 3 meses _____ ☐ 1

- De 3 a 6 meses _____ ☐ 2

- De 6 meses a 1 año _____ ☐ 3

- De 1 año a año y medio _____ ☐ 4

- De 1 año y medio a 2 años _____ ☐ 5

- 2 ó más años _____ ☐ 6

P43: ¿Cuándo empezaste a buscar trabajo?

- Antes de acabar la carrera de "... "___ ☐ 1
- Después de acabar la carrera de "... "___ ☐ 2

P44: ¿Durante cuánto tiempo estuviste buscando empleo?

- Menos de 3 meses_____ ☐ 1
- De 3 a 6 meses_____ ☐ 2
- De 6 meses a 1 año_____ ☐ 3
- De 1 año a año y medio _____ ☐ 4
- De 1 año y medio a 2 años___ ☐ 5
- 2 ó más años_____ ☐ 6

P45: ¿Cómo encontraste trabajo? (Marca tantas respuestas como sea necesario)

	Si marca esta opción	Si NO marca esta opción
A través de anuncios en el periódico, internet,...	1	2
A través de servicios de empleo públicos o de los servicios de empleo de la universidad	1	2
A través de empresas de trabajo temporal	1	2
Contacté con el empleador por iniciativa propia o utilicé contactos personales (familia, amigos)	1	2
El empleador se puso en contacto conmigo	1	2
Continué con las prácticas en empresas/ instituciones que realicé durante la carrera	1	2
Preparé unas oposiciones	1	2
Monté mi propio negocio	1	2
De otra forma	1	2

Ir a P52

P46: ¿Has buscado empleo desde que te titulaste?

- Sí_____ ☐ 1
- No_____ ☐ 2 → **ir a P51**

P47: ¿Cuándo empezaste a buscar trabajo?

- Antes de acabar la carrera de "... "___ ☐ 1
- Después de acabar la carrera de "... "___ ☐ 2

P48: ¿Cuánto tiempo llevas buscando empleo?

- Menos de 3 meses_____ ☐ 1
- De 3 a 6 meses_____ ☐ 2
- De 6 meses a 1 año_____ ☐ 3
- De 1 año a año y medio _____ ☐ 4
- De 1 año y medio a 2 años___ ☐ 5
- 2 ó más años_____ ☐ 6

P49: ¿Cómo buscas o has buscado trabajo? (Marca tantas respuestas como opciones hayas utilizado)

	Si marca esta opción	Si NO marca esta opción
A través de anuncios en el periódico, internet,...	1	2
A través de servicios públicos de empleo o de los servicios de empleo de la universidad	1	2
A través de empresas de trabajo temporal	1	2
Buscando ayuda entre profesionales conocidos del sector o contactos personales (familia, amigos)	1	2
Estoy preparando o prepararé unas oposiciones	1	2
Buscando equipamiento para establecer mi propio negocio	1	2
Otra forma	1	2

P50: Puntúa de 4 (muy importante) a 1 (nada importante) cada uno de los siguientes elementos en cuanto a tus dificultades para encontrar trabajo:

	Muy importante	Bastante importante	Poco importante	Nada importante
Carencias en la formación universitaria recibida	4	3	2	1

Actividades personales que impiden trabajar (seguir estudiando, familia, otras...)	4	3	2	1
Falta de experiencia profesional	4	3	2	1
Exigencia de tener un trabajo acorde con mis expectativas económicas, profesionales,...	4	3	2	1
Falta de conocimientos de idiomas	4	3	2	1
Falta de conocimientos informáticos	4	3	2	1

Ir a P52

P51: ¿Por cuál de los siguientes motivos no buscas empleo? (Marca tantas opciones como sea necesario)

	Si marca esta opción	Si NO marca esta opción
Enfermedad o discapacidad	1	2
Estar preparando oposiciones	1	2
Estar estudiando en este momento	1	2
Cuidado de los hijos, de otras personas dependientes u otras responsabilidades familiares	1	2
Jubilación (incluida jubilación anticipada)	1	2
Otros motivos	1	2

Ir a P85

P52: Desde que te titulaste, ¿has rechazado algún trabajo por considerarlo poco adecuado?

- Sí _____ ☐ 1
- No _____ ☐ 2

Si P52 = 2 y P36 = 1 → ir a P54

Si P52 = 2 y P40 = 1 → ir a P54

Si P52 = 2 y P41 = 1 → ir a P54

Si P52 = 2 y P46 = 1 → ir a P85

P53: ¿Por cuál/es de los siguientes motivos lo rechazaste?

	Si marca esta opción	Si NO marca esta opción

No me permitía continuar estudiando	1	2
El horario de trabajo era incompatible con obligaciones personales o familiares	1	2
El aceptarlo implicaba cambiar de lugar de residencia	1	2
El trayecto entre mi domicilio y el puesto de trabajo no estaba a una distancia razonable	1	2
Consideraba que no estaba adecuadamente remunerado	1	2
El nivel del puesto de trabajo no estaba acorde con mis habilidades y conocimientos	1	2
Otros motivos	1	2

Si P46 = 1 → ir a P85

2.4. Primer empleo posterior a su formación (Sólo se contesta si P36 = 1 o P40 = 1 o P41 = 1)

Las siguientes preguntas hacen referencia a la situación en la que estabas cuando empezaste a trabajar después de obtener el título de “...”.

Si seguiste más de seis meses en el trabajo que ya tenías antes de acabar la carrera, haz referencia a la situación en la que estabas inmediatamente después de titularte.

Excluye los trabajos esporádicos y no olvides incluir trabajos de prácticas, de formación, becas, ayuda en la empresa familiar...

P54: ¿En qué país o región desempeñabas este primer empleo?

- ESPAÑA _____ ☐ ESPA → ir a P55
- OTRO PAÍS ____ ☐ OTRO

P54.OTRO: Por favor, indica el país o región donde desempeñabas este primer empleo

País _____ ☐ Desplegable D4.2 → ir a P56

P55: ¿En qué provincia desempeñabas tu primer empleo?

Provincia _____ ☐ Desplegable D5

P56: ¿En cuál de las siguientes categorías se encontraba el puesto de trabajo que ocupabas después de titularte?

- Agrupaciones CNO-11 _____ ☐ Desplegable D2
- No sé _____ ☐ 99 → ir a P57

Si P56 = 01 → ir a P56.1

Si P56 = 02 → ir a P56.2

Si P56 = 03 → ir a P56.3

Si P56 = 04 → ir a P56.4

Si P56 = 05 → ir a P56.5

Si P56 = 06 → ir a P56.6

Si P56 = 07 → ir a P56.7

Si P56 = 08 → ir a P56.8

Si P56 = 09 → ir a P56.9 Si P56 = 20 → ir a P56.20 Si P56 = 10 → ir a P56.10 Si P56 = 21 → ir a P56.21

Si P56 = 11 → ir a P56.11

Si P56 = 12 → ir a P56.12

Si P56 = 13 → ir a P56.13

Si P56 = 14 → ir a P56.14

Si P56 = 15 → ir a P56.15

Si P56 = 16 → ir a P56.16

Si P56 = 17 → ir a P56.17

Si P56 = 18 → ir a P56.18

Si P56 = 19 → ir a P56.19

P56.1:

- Médico (cualquier especialidad), enfermero, matró, farmacéutico,
dentista, óptico, fisioterapeuta, podólogo, logopeda _____ ☐ 21
- Veterinario _____ ☐ 20
- _____
- Técnico sanitario de laboratorio, pruebas
diagnósticas, prótesis;
ayudante de veterinaria, higienista bucodental _____ ☐ 33
- Auxiliar de enfermería, celador, camillero, ayudante de farmacia ☐ 56
- No sé _____ ☐ 99 → ir a P57

Ir a P58**P56.2:** Por favor, indica el trabajo que desempeñabas

- Maestro, profesor, técnico de educación infantil _____ ☐ 22
- Profesor, técnico de educación especial; profesor de
clases particulares o academias _____ ☐ 23
- Deportista, entrenador profesional; monitor de deportes y
tiempo libre _____ ☐ 37
- Instructor de autoescuela _____ ☐ 58
- No sé _____ ☐ 99 → ir a P57

Ir a P58**P56.3:** Por favor, indica el trabajo que desempeñabas

- Abogado, fiscal, juez, notario, registrador, procurador ☐ 25
- Economista _____ ☐ 28
- _____
- Contable, analista o asesor financiero (excepto
tenedor
de libros, tasador, intermediario de cambio y bolsa...) ☐ 26
- Tenedor de libros, tasador, intermediario de cambio y bolsa, ... _____ ☐ 34
- Psicólogo, sociólogo, trabajador social _____ ☐ 28
- Bibliotecario, conservador, escritor, intérprete, traductor _____ ☐ 29
- Músico, pintor, escultor, actor/actriz, bailarín, torero _____ ☐ 29
- Periodista, locutor _____ ☐ 20
- No sé _____ ☐ 99 → ir a P57

Ir a P58

P56.4:

Programador informático; operador de sistemas, redes, web y asistencia al usuario; técnico en grabación audiovisual,

- radiodifusión, telecomunicaciones_____ ☐ 38
- Analista y diseñador de software, multimedia, bases de datos y redes informáticas_____ ☐ 27
 - Diseñador gráfico, multimedia_____ ☐ 24
 - Fotógrafo_____ ☐ 37
 - No sé_____ ☐ 99 → ir a P57

Ir a P58

P56.5: Por favor, indica el trabajo que desempeñabas

- Miembro del poder ejecutivo o de los cuerpos legislativos; directivo de la Administración Pública u organizaciones de interés social; director general, presidente ejecutivo_____ ☐ 11
- Director de producción u operaciones_____ ☐ 13
- Director de departamentos administrativos (recursos humanos, financieros,...) o comerciales (ventas, publicidad,...)_____ ☐ 12
- Director o gerente de empresas de alojamiento, restauración o comercio_____ ☐ 14
- Director o gerente de otras empresas de servicios (recreativas, deportivas, culturales, de gestión de residuos,...)_____ ☐ 15
- No sé_____ ☐ 99 → ir a P57

Ir a P58

P56.6: Por favor, indica el trabajo que desempeñabas

- Ingeniero técnico y superior (industrial, eléctrico, electrónico, de telecomunicaciones,...), arquitecto, aparejador, topógrafo, enólogo_____ ☐ 20
- Ingeniero técnico y superior (agronomo, de montes, ...)_____ ☐ 21
- Físico, químico, estadístico, meteorólogo, geógrafo, biólogo, zoólogo, botánico_____ ☐ 24
- Profesional en publicidad, marketing y relaciones públicas_____ ☐ 26
- Inspector o subinspector de Hacienda; otras ocupaciones

exclusivas de las AAPP (grupos A1 o A2)_____ ☐ 26

- No sé_____ ☐ 99 → ir a P57

Ir a P58

P56.7:

Oficial, suboficial, tropa o marinería de las Fuerzas Armadas_ ☐ 00

- Inspector o comisario de Policía, oficial de la Guardia Civil _____ ☐ 26
- Técnico de la Policía, suboficial de la Guardia Civil_____ ☐ 36
- Guardia civil, policía nacional, autonómico, municipal, vigilante privado, guardia jurado, vigilante forestal, vigilante de prisiones, bombero_____ ☐ 59 - Ambientólogo_____ ☐ 24
- No sé_____ ☐ 99 → ir a P57

Ir a P58

P56.8 Por favor, indica el trabajo que desempeñabas

- Técnico en electricidad, electrónica, construcción, laboratorio, ingenierías; técnico de control de procesos; técnico de control de calidad_____ ☐ 31
- Encargado o jefe de equipo de las industrias manufactureras_ ☐ 32
- No sé_____ ☐ 99 → ir a P57

Ir a P58

P56.9: Por favor, indica el trabajo que desempeñabas

- Visitador médico, representante de ventas técnicas_____ ☐ 26
- Agente o representante comercial; agente inmobiliario; agente de seguros, compras; consignatario_____ ☐ 35
- No sé_____ ☐ 99 → ir a P57

Ir a P58

P56.10: Por favor, indica el trabajo que desempeñabas

- Administrativo contable, de nóminas; auxiliar bancario; empleado administrativo de almacén, logística, transportes____ ☐ 41 - Auxiliar de justicia, oficial de notaría _____ ☐ 37
- Secretario o asistente administrativo, personal de expedición de licencias, otro profesional de apoyo de las AAPP (grupos C o D)_ ☐ 36
- Empleado de bibliotecas, archivos_____ ☐ 42

- Telefonista, teleoperador, recepcionista, empleado de agencias de viajes, empleado de información al usuario, cajero de bancos, empleado de mostrador de correos_____ ☐ 44
- Empleado de servicios de correos, codificador, corrector o servicios de personal_____ ☐ 42 - Grabador de datos_____ ☐ 43
- Otro personal administrativo con tareas de atención al público_ ☐ 45
- Otro personal administrativo sin tareas de atención al público_ ☐ 43
- No sé_____ ☐ 99 → ir a P57

Ir a P58

P56.11: Por favor, indica el trabajo que desempeñabas

- Camarero o cocinero propietario_____ ☐ 50
 - Camarero o cocinero asalariado_____ ☐ 51
 - Ayudante de cocina o preparador de comidas rápidas__ ☐ 93
 - No sé_____ ☐ 99
- ir a P57

Ir a P58

P56.12: Por favor, indica el trabajo que desempeñabas

- Dependiente_____ ☐ 52
 - Cajero (excepto bancos) y taquillero_____ ☐ 55
 - Vendedor en quioscos o en mercadillos, televendedor,
promotor comercial, modelo_____ ☐ 54
 - Vendedor de cupones, loterías, bingos_____ ☐ 44
 - Vendedor callejero fuera de mercadillos_____ ☐ 94
 - No sé_____ ☐ 99
- ir a P57

Ir a P58

P56.13: Por favor, indica el trabajo que desempeñabas

- Cuidador de niños (domicilios, guarderías), cuidador de otras
personas en domicilios____ ☐ 57 - Peluquero, esteticista____ ☐ 58

- Diseñador, decorador_____ ☐ 37
 - Delineante, diseñador técnico_____ ☐ 31
 - Azafato de avión, de tierra, guía turístico_____ ☐ 58
 - No sé_____ ☐ 99
- ir a P57

Ir a P58

P56.14: Por favor, indica el trabajo que desempeñabas

- Camionero, taxista, conductor de coche o furgoneta,
conductor de autobús ☐ 84 - Conductor de trenes ☐ 83

- Repartidor o mensajero en motocicleta ☐ 84

- No sé ☐ 99

→ ir a P57

Ir a P58

P56.15: Por favor, indica el trabajo que desempeñabas

- Empleado doméstico ☐ 91

- Personal de limpieza de oficinas, hoteles, limpiador de
vehículos, ventanas, otra limpieza a mano ☐ 92

- Conserje, portero de edificios ☐ 58

- Operador de máquinas de lavandería y tintorería ☐ 81

- Barrendero, recogedor de basura, clasificador de desechos ☐ 94

- No sé ☐ 99 → ir a P57

Ir a P58

P56.16: Por favor, indica el trabajo que desempeñabas

- Mecánico, ajustador de maquinaria ☐ 74

- Reparador de equipos eléctricos, electrónicos y de
telecomunicaciones ☐ 75

- Chapista, soldador, montador de estructuras metálicas, operador
de máquinas-herramienta, cerrajero, herrero, matricero ☐ 73

- No sé ☐ 99 → ir a P57

Ir a P58

P56.17: Por favor, indica el trabajo que desempeñabas

- Montador de productos industriales ☐ 82

- Operador de robots industriales ☐ 31

- Operador de maquinaria fija: horno, prensa, sierra, fresadora, máquinas
de embalaje, embotellado, etiquetado..., excepto artes
gráficas y control de procesos ☐ 81

- Oficial y operario de las artes gráficas ☐ 76

- Trabajador en cadena de producción, excepto AAGG y técnicos

- de control de procesos_____ ☐ 81
- No sé_____ ☐ 99 → ir a P57

Ir a P58

P56.18: Por favor, indica el trabajo que desempeñabas

- Elaborador de alimentos, bebidas o tabacos_____ ☐ 77
- Sastre, zapatero, bordador, tapicero, ebanista_____ ☐ 78
- Artesano (cerámica, vidrio, madera, cuero, textil), joyero, relojero ☐ 76
- No sé_____ ☐ 99 → ir a P57

Ir a P58

P56.19: Por favor, indica el trabajo que desempeñabas

- Jardinero, horticultor (huertas, viveros, invernaderos), agricultor_ ☐ 61
- Ganadero, pastor_____ ☐ 62
- Pescador, criador de especies acuáticas y trabajador forestal_ ☐ 64
- Conductor de tractor y otra maquinaria agrícola móvil_____ ☐ 83
- Trabajador agropecuario_____ ☐ 63
- Peón agrícola, de la ganadería o la pesca_____ ☐ 95
- No sé_____ ☐ 99
→ ir a P57

Ir a P58

P56.20: Por favor, indica el trabajo que desempeñabas

- Delineante, diseñador técnico_____ ☐ 31
- Jefe de obra, capataz, encargado_____ ☐ 32
- Albañil, encofrador, ferrallista, trabajador en hormigón, carpintero (madera, aluminio)_____ ☐ 71
- Conductor-operador de maquinaria pesada_____ ☐ 83
- Escayolista, yesista, pintor, empapelador, barnizador, parquetero, alicatador_____ ☐ 72

- Fontanero, instalador de refrigeración y climatización _____ ☐ 72
 - Electricista _____ ☐ 75
 - Instalador de material aislante, de placas de energía solar, cristalero
☐ 72
 - Minero, sondista _____ ☐ 81
 - Peón de la construcción o la minería _____ ☐ 96
 - No sé _____ ☐ 99
- ir a P57

Ir a P58

P56.21: Por favor, indica el trabajo que desempeñabas

- Reponedor de hipermercados, mozo de carga, mudanzas _____ ☐ 98
 - Peón industrial: lavador, embotellador... a mano _____ ☐ 97
 - Repartidor (publicidad, prensa gratuita, etc), ordenanza y mozo
de equipaje _____ ☐ 94
 - Lector de contadores y recaudador de máquinas expendedoras _____ ☐
94
 - No sé _____ ☐ 99
- ir a P57

Ir a P58

P57: Por favor, escribe cuál era el puesto de trabajo en tu primer empleo después de obtener el título de "..."

- Literal de ocupación:
.....

P58: ¿Cuál de las siguientes era tu situación profesional?

- Trabajador en prácticas, formación o becario _____ ☐ 1
- Asalariado con trabajo permanente o contrato de trabajo indefinido ☐ 2
- Asalariado con trabajo temporal o contrato de trabajo de duración

- determinada_____ ☐ 3
- Empresario o trabajador independiente ☐ 4
 - Ayuda en la empresa o negocio familiar ☐ 5

P59: ¿Cómo era tu jornada laboral?

- A tiempo completo _____ ☐ 1
- A tiempo parcial _____ ☐ 2

P60: Por favor, indica en cuál de los siguientes intervalos se encontraba tu sueldo mensual neto

- Menos de 650 euros _____ ☐ 1
- De 650 a menos de 1300 _____ euros ☐ 2
- De 1300 a menos de 2000 _____ euros ☐ 3
- De 2000 a menos de 2650 _____ euros ☐ 4 -
- De 2650 a menos de _____ 3300 euros ☐ 5
- De 3.300 euros en adelante _____ ☐ 6
- No sé _____ ☐ 9

P61: ¿Cuál crees que era el nivel de formación más apropiado para realizar ese trabajo?

- Doctorado o postdoctorado _____ ☐ 1
- Título universitario (excepto doctorado o postdoctorado) ☐ 2
- Formación profesional de grado superior _____ ☐ 3
- Formación profesional de grado medio/bachillerato _____ ☐ 4
- ESO, EGB, etc. _____ ☐ 5

P62: ¿Cuál era el área de estudio más apropiada para ese trabajo?

- Exclusivamente el área de estudios del título de "...".__ ☐ 1
- El área de estudios del título de "..." o algún área relacionada_ ☐ 2

- Un área totalmente diferente_____ ☐ 3
- Ningún área en particular _____ ☐ 4

P63: ¿Se hacía uso en ese trabajo de los conocimientos y habilidades adquiridos en esta carrera?

- Sí___ ☐ 1
- No_____ ☐ 2

P64: ¿Continúas en este primer trabajo?

- Sí___ ☐ 1 → Ir a P74
- No_____ ☐ 2

2.5. Características de su empleo principal actual

Por favor, responde ahora a estas preguntas sobre tu situación laboral actual.

P65: ¿En qué situación laboral te encuentras actualmente? (Sólo se contesta si P64 = 2)

- Desempeño un trabajo o profesión (incluido el trabajo no remunerado para una empresa o negocio familiar y prácticas o becas de formación remuneradas) a tiempo completo _____ ☐ 01 → Ir a P68
- Desempeño un trabajo o profesión (incluido el trabajo no remunerado para una empresa o negocio familiar y prácticas o becas de formación remuneradas) a tiempo parcial _____ ☐ 02 → Ir a P68
- Desempleado_____ ☐ 03
- Estudiante (incluida experiencia laboral no remunerada)_____ ☐ 04
- Jubilado (incluido jubilado anticipadamente)_____ ☐ 05
- Incapacitado permanentemente para trabajar_____ ☐ 06
- Dedicado a las labores del hogar, cuidado de niños u otras personas ☐ 07
- Otra situación_____ ☐ 08

P66: Durante las 4 últimas semanas ¿has buscado trabajo activamente? (Has buscado trabajo activamente si te has puesto en contacto con una oficina de empleo de la Administración o privada, con empresarios, familiares, amigos, un sindicato...

O si has puesto, consultado o contestado anuncios en prensa, radio, televisión, internet...

O si has realizado un examen o una entrevista.

O si has buscado terrenos, un local o maquinaria, financiación, o solicitado permisos o licencias para establecerte por tu cuenta.

No se considera búsqueda activa si estás a la espera de resultados de solicitudes anteriores, de la llamada de una oficina de empleo de la Administración o de los resultados de una oposición) .

- Sí ☐ 1
- No ☐ 2 → Ir a P77

P67: Si encontraras un trabajo adecuado en este momento, ¿estarías disponible para empezar a trabajar?

- Sí ☐ 1
- No ☐ 2

Ir a P77

P68: ¿En qué país o región desempeñas tu labor profesional?

- ESPAÑA ☐ ESPA → Ir a P69
- OTRO PAÍS ☐ OTRO

P68.OTRO: Por favor, indica el país o región donde desempeñas tu labor profesional

- País ☐ Desplegable D4.2 → Ir a P70

P69: ¿En qué provincia desempeñas tu labor profesional?

- Provincia ☐ Desplegable D5

P70: ¿Cuál de las siguientes es tu situación profesional?

- Trabajador en prácticas, formación o becario _____ ☐ 1
- Asalariado con trabajo permanente o contrato _____ ☐ 2
- Asalariado con trabajo temporal o contrato de trabajo de duración determinada _____ ☐ 3
- Empresario o trabajador independiente _____ ☐ 4
- Ayuda en la empresa o negocio familiar _____ ☐ 5
- No sé _____ ☐ 9

P71: En tu empleo principal actual, ¿en cuál de las siguientes categorías se encuentra tu puesto de trabajo?

- Agrupaciones CNO-11 _____ ☐ Desplegable D2
- No sé _____ ☐ 99 → ir a P73

Si P71 = 01 → ir a P72.1	Si P71 = 12 → ir a P72.12
Si P71 = 02 → ir a P72.2	Si P71 = 13 → ir a P72.13
Si P71 = 03 → ir a P72.3	Si P71 = 14 → ir a P72.14
Si P71 = 04 → ir a P72.4	Si P71 = 15 → ir a P72.15
Si P71 = 05 → ir a P72.5	Si P71 = 16 → ir a P72.16
Si P71 = 06 → ir a P72.6	Si P71 = 17 → ir a P72.17
Si P71 = 07 → ir a P72.7	Si P71 = 18 → ir a P72.18
Si P71 = 08 → ir a P72.8	Si P71 = 19 → ir a P72.19
Si P71 = 09 → ir a P72.9	Si P71 = 20 → ir a P72.20
Si P71 = 10 → ir a P72.10	Si P71 = 21 → ir a P72.21
Si P71 = 11 → ir a P72.11	

P72.1: Por favor, indica el trabajo que desempeñas

- Médico (cualquier especialidad), enfermero, matró, farmacéutico, dentista, óptico, fisioterapeuta, podólogo, logopeda _____ ☐ 21
- Veterinario _____ ☐ 20
- Técnico sanitario de laboratorio, pruebas diagnósticas y prótesis,

- ayudante de veterinaria, higienista bucodental_____ ☐ 33
- Auxiliar de enfermería, celador, camillero, ayudante de farmacia ☐ 56
- No sé_____ ☐ 99 → ir a P73

Ir a P74

P72.2: Por favor, indica el trabajo que desempeñas

- Maestro y profesor, y técnico de educación infantil_____ ☐ 22
- Profesor y técnico de educación especial, profesor de clases particulares y academias_____ ☐ 23
- Deportista y entrenador profesional, monitor de deportes y tiempo libre_____ ☐ 37
- Instructor de autoescuela_____ ☐ 58
- No sé_____ ☐ 99 → ir a P73

Ir a P74

P72.3: Por favor, indica el trabajo que desempeñas

- Abogado, fiscal, juez, notario, registrador, procurador_____ ☐ 25
- Economista_____ ☐ 28
- Contable, analista y asesor financiero (excepto tenedor de libros, tasador, intermediario de cambio y bolsa...)_____ ☐ 26
- Tenedor de libros, tasador, intermediario de cambio y bolsa _____ ☐ 34
- Psicólogo, sociólogo, trabajador social_____ ☐ 28
- Bibliotecario, conservador, escritor, intérprete, traductor_____ ☐ 29
- Músico, pintor, escultor, actor/actriz, bailarín, torero_____ ☐ 29
- Periodista, locutor_____ ☐ 20
- No sé_____ ☐ 99
→ ir a P73

Ir a P74

P72.4: Por favor, indica el trabajo que desempeñas

- Programador informático, operador de sistemas, redes, web y asistencia al usuario y técnico en grabación audiovisual, radiodifusión y telecomunicaciones_____ ☐ 38
- Analista y diseñador de software, multimedia, bases de datos y redes informáticas_____ ☐ 27
- Diseñador gráfico y multimedia_____ ☐ 24
- Fotógrafo_____ ☐ 37
- No sé_____ ☐ 99

→ ir a P73

Ir a P74

P72.5: Por favor, indica el trabajo que desempeñas

- Miembro del poder ejecutivo y de los cuerpos legislativos; directivo de la Administración Pública y organizaciones de interés social; director general y presidente ejecutivo_____ ☐ 11 -

Director de producción y operaciones_____

☐ 13

- Director de departamentos administrativos (recursos humanos, financieros,...) y comerciales (ventas, publicidad,...)_____ ☐ 12
- Director y gerente de empresas de alojamiento, restauración y comercio_____ ☐ 14
- Director y gerente de otras empresas de servicios (recreativas, deportivas, culturales, de gestión de residuos,...)_____ ☐ 15
- No sé_____ ☐ 99

→ ir a P73

Ir a P74

P72.6: Por favor, indica el trabajo que desempeñas

- Ingeniero técnico y superior (industrial, eléctrico, electrónico, de telecomunicaciones,...), arquitecto, aparejador, topógrafo, enólogo_____ ☐ 20
- Ingeniero técnico y superior (agronomo, de montes, ...)_____ ☐ 21
- Físico, químico, estadístico, meteorólogo, geógrafo, biólogo, zoólogo, botánico_____ ☐ 24

- Profesional en publicidad, marketing y relaciones públicas_____ ☐ 26
 - Inspector y subinspector de Hacienda, otras ocupaciones exclusivas de las AAPP (grupos A1 o A2)_____ ☐ 26
 - No sé_____ ☐ 99
- ir a P73

Ir a P74

P72.7: Por favor, indica el trabajo que desempeñas

- Oficial, suboficial, tropa y marinería de las Fuerzas Armadas_ ☐ 00 - Inspector y comisario de Policía, oficial de la Guardia Civil _____ ☐ 26
 - Técnico de la Policía, suboficial de la Guardia Civil_____ ☐ 36
 - Guardia civil, policía nacional, autonómico, municipal, vigilante privado, guardia jurado, vigilante forestal, vigilante de prisiones, bombero_____ ☐ 59
 - Ambientólogo_____ ☐ 24
 - No sé_____ ☐ 99
- ir a P73

Ir a P74

P72.8: Por favor, indica el trabajo que desempeñas

- Técnico en electricidad, electrónica, construcción, laboratorio e ingenierías, técnico de control de procesos y técnico de control de calidad_____ ☐ 31
 - Encargado y jefe de equipo de las industrias manufactureras_ ☐ 32
 - No sé_____ ☐ 99
- ir a P73

Ir a P74

P72.9: Por favor, indica el trabajo que desempeñas

- Visitador médico, representante de ventas técnicas_____ ☐ 26
- Agente y representante comercial, agente inmobiliario, agente de seguros, compras, consignatario_____ ☐ 35

- No sé_____ ☐ 99
→ ir a P73

Ir a P74

P72.10: Por favor, indica el trabajo que desempeñas

- Administrativo contable, de nóminas, auxiliar bancario, empleado administrativo de almacén, logística y transportes_____ ☐ 41
- Auxiliar de justicia, oficial de notaría_____ ☐ 37
- Secretario y asistente administrativo, personal de expedición de licencias, otro profesional de apoyo de las AAPP (grupos C o D)_____ ☐ 36
- Empleado de bibliotecas, archivos_____ ☐ 42
- Telefonista, teleoperador, recepcionista, empleado de agencias de viajes, empleado de información al usuario, cajero de bancos, empleado de mostrador de correos_____ ☐ 44
- Empleado de servicios de correos, codificador, corrector y servicios de personal_____ ☐ 42 -
- Grabador de datos_____ ☐ 43
- Otro personal administrativo con tareas de atención al público_____ ☐ 45
- Otro personal administrativo sin tareas de atención al público_____ ☐ 43
- No sé_____ ☐ 99
→ ir a P73

Ir a P74

P72.11: Por favor, indica el trabajo que desempeñas

- Camarero y cocinero propietario_____ ☐ 50
- Camarero y cocinero asalariado_____ ☐ 51
- Ayudante de cocina y preparador de comidas rápidas_____ ☐ 93
- No sé_____ ☐ 99
→ ir a P73

Ir a P74

P72.12: Por favor, indica el trabajo que desempeñas

- Dependiente_____ ☐ 52
- Cajero (excepto bancos) y taquillero_____ ☐ 55
- Vendedor en quioscos o en mercadillos, televendedor,
promotor comercial, modelo _____ ☐ 54
- Vendedor de cupones, loterías, bingos_____ ☐ 44
- Vendedor callejero fuera de mercadillos_____ ☐ 94
- No sé_____ ☐ 99

→ ir a P73

Ir a P74

P72.13: Por favor, indica el trabajo que desempeñas

- Cuidador de niños (domicilios, guarderías) y cuidador de
otras personas en domicilios _____ ☐ 57
- Peluquero, esteticista_____ ☐ 58
- Diseñador, decorador_____ ☐ 37
- Delineante, diseñador técnico_____ ☐ 31
- Azafato de avión, de tierra, guía turístico_____ ☐ 58
- No sé_____ ☐ 99 → ir a P73

Ir a P74

P72.14: Por favor, indica el trabajo que desempeñas

- Camionero, taxista, conductor de coche o furgoneta, conductor
de autobús_____ ☐ 84
- Conductor de trenes_____ ☐ 83
- Repartidor y mensajero en motocicleta_____ ☐ 84
- No sé_____ ☐ 99 → ir a P73

Ir a P74

P72.15: Por favor, indica el trabajo que desempeñas

- Empleado doméstico_____ ☐ 91
- Personal de limpieza de oficinas, hoteles, limpiador de vehículos, ventanas, otra limpieza a mano_____ ☐ 92
- Conserje, portero de edificios_____ ☐ 58
- Operador de máquinas de lavandería y tintorería_____ ☐ 81
- Barrendero, recogedor de basura, clasificador de desechos_____ ☐ 94
- No sé_____ ☐ 99 → ir a P73

Ir a P74

P72.16: Por favor, indica el trabajo que desempeñas

- Mecánico, ajustador de maquinaria_____ ☐ 74
- Reparador de equipos eléctricos, electrónicos y de telecomunicaciones_____ ☐ 75
- Chapista, soldador, montador de estructuras metálicas, operador de máquinas-herramienta, cerrajero, herrero, matricero_____ ☐ 73
- No sé_____ ☐ 99
→ ir a P73

Ir a P74

P72.17: Por favor, indica el trabajo que desempeñas

- Montador de productos industriales_____ ☐ 82
- Operador de robots industriales_____ ☐ 31
- Operador de maquinaria fija: horno, prensa, sierra, fresadora, máquinas de embalaje, embotellado, etiquetado..., excepto artes gráficas y control de procesos_____ ☐ 81 - Oficial y operario de las artes gráficas_____ ☐ 76
- Trabajador en cadena de producción, excepto AAGG y técnicos de control de procesos_____ ☐ 81
- No sé_____ ☐ 99
→ ir a P73

Ir a P74

P72.18: Por favor, indica el trabajo que desempeñas

- Elaborador de alimentos, bebidas o tabacos_____ ☐ 77
 - Sastre, zapatero, bordador, tapicero, ebanista_____ ☐ 78
 - Artesano (cerámica, vidrio, madera, cuero, textil), joyero, relojero_ ☐ 76
 - No sé_____ ☐ 99
- ir a P73

Ir a P74

P72.19: Por favor, indica el trabajo que desempeñas

- Jardinero, horticultor (huertas, viveros, invernaderos), agricultor_ ☐ 61
 - Ganadero, pastor_____ ☐ 62
 - Pescador, criador de especies acuáticas y trabajador forestal_ ☐ 64
 - Conductor de tractor y otra maquinaria agrícola móvil _____ ☐ 83
 - Trabajador agropecuario_____ ☐ 63
 - Peón agrícola, de la ganadería o la pesca_____ ☐ 95
 - No sé_____ ☐ 99
- ir a P73

Ir a P74

P72.20: Por favor, indica el trabajo que desempeñas

- Delineante, diseñador técnico_____ ☐ 31
- Jefe de obra, capataz, encargado_____ ☐ 32
- Albañil, encofrador, ferrallista, trabajador en hormigón, carpintero (madera, aluminio)_____ ☐ 71 - Conductor-operador de maquinaria pesada_ ☐ 83
- Escayolista, yesista, pintor, empapelador, barnizador, parquetero, alicatador_____ ☐ 72

- Fontanero, instalador de refrigeración y climatización _____ ☐ 72
 - Electricista _____ ☐ 75
 - Instalador de material aislante, de placas de energía solar, cristalero ☐ 72
 - Minero, sondista _____ ☐ 81
 - Peón de la construcción o la minería _____ ☐ 96
 - No sé _____ ☐ 99
- ir a P73

Ir a P74

P72.21: Por favor, indica el trabajo que desempeñas

- Reponedor de hipermercados, mozo de carga, mudanzas _____ ☐ 98
 - Peón industrial: lavador, embotellador... a mano _____ ☐ 97
 - Repartidor (publicidad, prensa gratuita, etc.), ordenanza y mozo de equipaje _____ ☐ 94
 - Lector de contadores y recaudador de máquinas expendedoras _____ ☐ 94
 - No sé _____ ☐ 99
- ir a P73

Ir a P74

P73: Por favor, escribe cuál es tu principal puesto de trabajo actual

- Literal de ocupación:
.....

P74: ¿Cuál es la actividad principal del establecimiento o local donde trabajas? (Sólo se pregunta si

no está en los ficheros de la Tesorería de la Seguridad Social y (o P64 = 1 o P65 = (01 o 02)))

- CNAE 2009 (2 dígitos) _____ ☐ Desplegable D3
- No sé _____ ☐ 99

P75: ¿Cuántas personas trabajan en el mismo establecimiento o local donde trabajas? (Sólo se pregunta si no está en los ficheros de la Tesorería de la Seguridad Social y (o P64 = 1 o P70 = (1, 2, 3 o 5)))

- De 1 a 10 personas _____ ☐ 1
- Entre 11 - 19 personas _____ ☐ 2 - Entre 20 - 49 personas _____ ☐ 3
- 50 o más personas _____ ☐ 4
- No sé pero trabajan más de 10 personas _____ ☐ 5
- No sé _____ ☐ 9

P76: ¿En qué año comenzaste a trabajar para la empresa o negocio actual? (Sólo se pregunta si no está en los ficheros de la Tesorería de la Seguridad Social y (o P64 = 1 o P65 = (01 o 02)))

- Año _____

2.6. Historial laboral

Ahora vamos a hacerte unas preguntas que resumen tu experiencia laboral desde que terminaste los estudios de "..."

P77: ¿Para cuántos empleadores distintos has trabajado desde que obtuviste el título de "..."? (Sólo se pregunta si no está en los ficheros de la Tesorería de la Seguridad Social y (o P36 = 1 o P40 = 1 o P41 = 1))

- Número de empleadores _____

P78: Aproximadamente, ¿cuánto tiempo has estado trabajando en total desde que obtuviste el título de "..."? (Sólo se pregunta si no está en los ficheros de la Tesorería de la Seguridad Social y (o P36 = 1 o P40 = 1 o P41 = 1))

- Menos de 6 meses_____ ☐ 1
- De 6 meses a 1 año_____ ☐ 2
- De 1 año a año y medio___ ☐ 3
- De 1 año y medio a 2 años _____ ☐ 4
- 2 ó más años_____ ☐ 5

P79: ¿Cuántas veces has estado en paro desde que obtuviste el título de "..."? (Sólo se pregunta si no está en los ficheros de la Tesorería de la Seguridad Social y (o P36 = 1 o P40 = 1 o P41 = 1))

- Veces en paro _____ ☐ ☐

P80: ¿Has intentado conseguir otro trabajo remunerado en las últimas 4 semanas? (Si estás preparando oposiciones responde **SÍ** a esta pregunta) (Sólo se pregunta si P66 ≠ (1 y 2) y (o P36 = 1 o P40 = 1 o P41 = 1))

- Sí _____ ☐ 1
- No, pero estoy esperando respuesta a solicitudes anteriores de empleo _____ ☐ 2
- No _____ ☐ 3

3. DISTANCIA ENTRE EMPRESA Y EMPLEADO

A continuación te vamos a hacer unas preguntas para conocer las diferencias entre tu nivel de formación y el nivel de formación más adecuado a tu puesto de trabajo, y los factores que han influido en que consiguieras un empleo.

P81: Teniendo en cuenta tu TRABAJO ACTUAL (el principal) o tu ÚLTIMO TRABAJO, y según tu opinión ¿cuál es, o era, el nivel de formación más apropiado para realizar este trabajo? (Sólo se pregunta si P64 = 2 y (o P36 = 1 o P40 = 1 o P41 = 1))

- Doctorado o postdoctorado_____ ☐ 1
- Título universitario (excepto doctorado o postdoctorado) ☐ 2
- Formación profesional de grado superior _____ ☐ 3
- Formación profesional de grado medio/ bachillerato_____ ☐ 4
- ESO, EGB, etc. _____ ☐ 5

P82: Teniendo en cuenta tu TRABAJO ACTUAL (el principal) o tu ÚLTIMO TRABAJO, ¿cuál crees que es, o era, el área de estudio más apropiada para este trabajo? (Sólo se pregunta si P64 = 2 y (o P36 = 1 o P40 = 1 o P41 = 1))

- Exclusivamente el área de estudios del título de "...". ☐ 1
- El área de estudios del título de "..." o algún área relacionada ☐ 2
- Un área totalmente diferente ☐ 3
- Ningún área en particular ☐ 4

P83: Para conseguir este trabajo, valora de 5 (mucho) a 1 (nada) cuánto han influido los siguientes factores: (Ten en cuenta tu TRABAJO ACTUAL (el principal) o tu ÚLTIMO TRABAJO) (Sólo se pregunta si (P65 = (01 o 02) y P70 ≠ (4 y 5)) o (P58 ≠ (4 y 5) y P64 = 1))

	Mucho	Bastante	Algo	Poco	Nada
Los conocimientos teóricos	5	4	3	2	1
Los conocimientos prácticos	5	4	3	2	1
Saber idiomas	5	4	3	2	1
La formación o el dominio de la informática y de las nuevas tecnologías	5	4	3	2	1
La manera de ser: personalidad, habilidades sociales, comunicación, capacidad de trabajar en grupo	5	4	3	2	1
La capacidad de gestión y planificación	5	4	3	2	1

P84: ¿Crees que te ha servido tu título de "..." para encontrar trabajo? (Sólo se pregunta si P36 = 1 o

P40 = 1 o P41 = 1)

- Sí ☐ 1
- No ☐ 2

4. MOVILIDAD

Ahora necesitamos que respondas a una serie de preguntas sobre tu movilidad geográfica. Comenzaremos con preguntas sobre la movilidad dentro de España y continuaremos con preguntas sobre tus desplazamientos a otros países.

Recuerda que estas preguntas hacen referencia a la movilidad geográfica tras obtener el título de "...".

4.1. Movilidad nacional

P85: ¿En qué país resides actualmente?

- ESPAÑA _____ ☐ ESPA → **ir a P86**
- OTRO PAÍS _____ ☐ OTRO

P85.OTRO: Por favor, indica el país o región donde resides actualmente

- País o región _____ ☐ Desplegable D4.2 → **ir a P90**

P86: Después de titularte, ¿te has ido a vivir a otra provincia dentro de España?

- Sí _____ ☐ 1
- No _____ ☐ 2 → **ir a P89**

P87: ¿A qué provincia te trasladaste?

- Provincia _____ ☐ Desplegable D5

P88: ¿Cuál/es de las siguientes razones te han llevado a tomar esa decisión?

	Si marca esta opción	Si NO marca esta opción
Haber encontrado un trabajo	1	2
Buscar empleo u otros motivos económicos	1	2

Factores académicos: continuar estudiando o investigando	1	2
Factores personales o familiares	1	2
Otros motivos	1	2

4.2. Movilidad internacional

A continuación, por favor responde a una serie de preguntas sobre tu movilidad fuera de España.

Recuerda que estas preguntas hacen referencia a la movilidad hacia el extranjero tras obtener el título de

"...".

P89: ¿Has vivido fuera de España después de titularte?

- Sí ☐ 1 → ir a P91
- No ☐ 2 → ir a P93

P90: Aparte del país en el que resides actualmente, ¿has vivido en algún otro país distinto de España

después de obtener la titulación de "...".

- Sí ☐ 1
- No ☐ 2 → ir a P92

P91: Por favor, dime en qué región/es o país/es distinto/s de España has vivido desde que te titulaste.

PAÍS 1 ☐ Desplegable D4.2 PAÍS 2 ☐

Desplegable D4.2

PAÍS 3 ☐ Desplegable D4.2

P92: ¿Cuál/es fueron los motivos que te llevaron a vivir fuera de España?

	Si marca esta opción	Si NO marca esta opción
Haber encontrado un trabajo	1	2
Buscar empleo u otros motivos económicos	1	2
Factores académicos: continuar estudiando o investigando	1	2
Factores personales o familiares	1	2
Factores políticos o administrativos (finalización de permiso de residencia o visado...)	1	2
Otros motivos	1	2

Si P85 ≠ ESPA → Ir a P96

P93: ¿Tienes previsto o te has planteado marcharte a vivir fuera de España? (Sólo se pregunta si P85 = ESPA)

- Sí _____ ☐ 1
- No _____ ☐ 2 → ir a P97
- No sé _____ ☐ 9 → ir a P97

P94: ¿Cuál/es de la siguientes razones te han llevado o te llevarían a tomar esa decisión?

	Si marca esta opción	Si NO marca esta opción
Haber encontrado un trabajo	1	2
Buscar empleo u otros motivos económicos	1	2
Factores académicos: continuar estudiando o investigando	1	2
Factores personales o familiares	1	2
Factores políticos o administrativos (finalización de permiso de residencia o visado...)	1	2
Otros motivos	1	2

P95: ¿A qué país o región tienes previsto o tendrías previsto trasladarte?

- País _____ ☐ Desplegable D4.2
- No sé _____ ☐ 9999

Ir a P97

P96: ¿Tienes previsto volver a vivir a España?

- Sí _____ ☐ 1
- No _____ ☐ 2

7. INFORMACIÓN DEL HOGAR

Por favor, indica los siguientes datos relativos al hogar en el que resides

P97: ¿De qué tipo es tu hogar?

- Vivo solo _____ ☐ 1 → ir a P99
- Convive un padre o una madre solo/a con algún hijo _____ ☐ 2
- Vive una pareja con hijos que conviven en el hogar _____ ☐ 3
- Vive una pareja sin hijos o una pareja con hijos que no conviven en el hogar _____ ☐ 4 → ir a P99
- Conviven una pareja, o un padre o una madre, con hijos y otras personas _____ ☐ 5
- Otro tipo de hogar _____ ☐ 6 → ir a P99

P98: ¿Es alguno de estos hijos menor de 25 años?

- Sí _____ ☐ 1

- No _____ ☐ 2

P99: ¿Tienes hijos?

- Sí _____ ☐ 1
- No _____ ☐ 2

Si P97 = 1 → Ir a P102

P100: ¿Con cuántas personas vives y compartes gastos y/o ingresos? (Si no compartes presupuesto marca 0)

- Número de personas _____ ☐ ☐ 0-12
- No sé _____ ☐ ☐ 99 → ir a P102

P101: Por favor, indica el número de personas del hogar de 16-64 años que están trabajando actualmente (sin incluirte a ti) (Sólo se pregunta si P100 ≠ (0 y 99))

- Número de personas del hogar de 16-64 años empleadas _____ ☐ ☐ 0-12
- No sé _____ ☐ ☐ 99

P102: Por favor, indica un teléfono de contacto

- Teléfono _____ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

P103: Si deseas recibir un resumen de los resultados, por favor escribe tu correo electrónico a continuación:

Mi correo electrónico es:

DESPLEGABLES

D1 (Utilizado en P25 y P27)

CÓDIGO	IDIOMA
SPA	Español
OSP	Idioma cooficial (gallego, catalán, euskera o valenciano)
ENG	Inglés
FRE	Francés
GER	Alemán
ITA	Italiano
POR	Portugués
RUS	Ruso
ARA	Árabe
CHI	Chino
OTH	Otros idiomas

D2 (Utilizado en P56 y P71)

CODIGO_GRUPO	TITULO_GRUPO
01	Personal sanitario
02	Personal docente y profesionales del deporte
03	Profesionales del derecho, la economía, las ciencias sociales y la cultura
04	Tecnologías de la información, las comunicaciones y audiovisuales
05	Directores y gerentes
06	Otras ocupaciones propias de estudios universitarios (ingeniero, arquitecto, matemático, físico, biólogo, profesional de la publicidad, grupos a1 y a2 de las AAPP ...)
07	Seguridad, defensa y protección civil
08	Técnicos y supervisores de la industria
09	Representantes, agentes comerciales
10	Personal administrativo
11	Cocineros, camareros
12	Dependientes, vendedores, cajeros

13	Servicios personales y otros servicios (cuidador, peluquero, decorador, delineante, azafato...)
14	Camioneros, taxistas, mensajeros y otros conductores
15	Servicio de limpieza, conserjes
16	Mecánicos, reparadores, soldadores, operadores de máquinas herramienta
17	Obreros en producción mecanizada, montadores
18	Trabajadores de la industria tradicional, artesanos
19	Agricultores, ganaderos, pescadores y sus peones
20	Trabajadores de la construcción o la minería
21	Peones y trabajadores manuales de industrias y servicios

D3 (Utilizado en P74)

CODIGO CNAE2009	TITULO_CNAE2009
AA	Agricultura, ganadería, silvicultura y pesca
BB	Industrias extractivas
CC	Industria manufacturera
DD	Suministro de energía eléctrica, gas, vapor y aire acondicionado
EE	Suministro de agua, actividades de saneamiento, gestión de residuos y descontaminación
FF	Construcción e ingeniería civil
GG	Comercio al por mayor y al por menor; reparación de vehículos de motor y motocicletas
HH	Transporte, almacenamiento, actividades postales y de correos
II	Hostelería
JJ	Información y comunicaciones (actividades de edición, cinematográficas, de vídeo, de sonido, programas de televisión, telecomunicaciones, programación informática y servicios de información)
KK	Actividades financieras y de seguros
LL	Actividades inmobiliarias
M1	Actividades profesionales, científicas y técnicas
M2	Actividades de fotografía
M3	Actividades veterinarias
N1	Actividades de alquiler
N2	Actividades relacionadas con el empleo

N3	Actividades de agencias de viajes, operadores turísticos, servicios de reservas y actividades relacionadas con los mismos
N4	Actividades de seguridad e investigación
N5	Servicios a edificios y actividades de jardinería
N6	Actividades administrativas de oficina y otras actividades auxiliares a las empresas
OO	Administración Pública y defensa; Seguridad Social obligatoria
PP	Educación
QQ	Actividades sanitarias y de servicios sociales
RR	Actividades artísticas, culturales, recreativas, deportivas y de entretenimiento
S1	Actividades asociativas y sindicales
S2	Reparación de ordenadores
S3	Efectos personales y artículos de uso doméstico y otros servicios personales (lavado y limpieza de prendas, peluquería y otros tratamientos de belleza, pompas fúnebres y actividades relacionadas)
S4	Actividades de mantenimiento físico.
TT	Particulares como empleadores de personal doméstico; particulares como productores de bienes y servicios para uso propio
UU	Actividades de organizaciones y organismos extraterritoriales

D4.1 (Utilizado en P8)

CÓDIGO	PAÍS
ITAL	ITALIA
PORT	PORTUGAL
ALEM	ALEMANIA
FRAN	FRANCIA
RUNI	REINO UNIDO
PNOR	PAÍSES NÓRDICOS (SUECIA, NORUEGA, FINLANDIA, DINAMARCA e ISLANDIA)
EU28	OTRO PAÍS DE LA UNIÓN EUROPEA
OTEU	OTRO PAÍS EUROPEO
ANOR	AMÉRICA DEL NORTE
ACYS	AMÉRICA CENTRAL Y DEL SUR

ASIA	ASIA
AFRI	ÁFRICA
OCEA	OCEANÍA

D4.2 (Utilizado en P54.OTRO, P68.OTRO, P85.OTRO y P91)

CÓDIGO	PAÍS
RUNI	REINO UNIDO
ALEM	ALEMANIA
FRAN	FRANCIA
EU28	OTRO PAÍS DE LA UNIÓN EUROPEA
OTEU	OTRO PAÍS EUROPEO
ANOR	AMÉRICA DEL NORTE
ACYS	AMÉRICA CENTRAL Y DEL SUR
ASIA	ASIA
OCEA	OCEANÍA
AFRI	ÁFRICA

D5 (Utilizado en P55, P69 y P87)

CÓDIGO	PROVINCIA	CÓDIGO	PROVINCIA
01	Araba/Álava	27	Lugo
02	Albacete	28	Madrid
03	Alicante/Alacant	29	Málaga
04	Almería	30	Murcia
05	Ávila	31	Navarra
06	Badajoz	32	Ourense
07	Balears, Illes	33	Asturias
08	Barcelona	34	Palencia
09	Burgos	35	Palmas, Las
10	Cáceres	36	Pontevedra
11	Cádiz	37	Salamanca
12	Castellón/Castelló	38	Santa Cruz de Tenerife
13	Ciudad Real	39	Cantabria
14	Córdoba	40	Segovia
15	Coruña, A	41	Sevilla
16	Cuenca	42	Soria
17	Girona	43	Tarragona
18	Granada	44	Teruel
19	Guadalajara	45	Toledo
20	Gipuzkoa	46	Valencia/València
21	Huelva	47	Valladolid
22	Huesca	48	Bizkaia
23	Jaén	49	Zamora
24	León	50	Zaragoza
25	Lleida	51	Ceuta
26	Rioja, La	52	Melilla

D6 (Utilizado en P18)

CÓDIGO	DENOMINACIÓN DE LA UNIVERSIDAD
001	Universidad de Alicante
002	Universidad de Extremadura
003	Universitat de les Illes Balears
004	Universitat de Barcelona
005	Universidad de Cádiz
006	Universidad de Córdoba
007	Universidade de Santiago de Compostela
008	Universidad de Granada
009	Universidad de León
010	Universidad Complutense de Madrid
011	Universidad de Málaga
012	Universidad de Murcia
013	Universidad de Oviedo
014	Universidad de Salamanca
015	Universidad de La Laguna
016	Universidad de Cantabria
017	Universidad de Sevilla
018	Universitat de València-Estudi General
019	Universidad de Valladolid
020	Universidad del País Vasco/Euskal Herriko Unibertsitatea
021	Universidad de Zaragoza
022	Universitat Autònoma de Barcelona
023	Universidad Autónoma de Madrid
024	Universitat Politècnica de Catalunya
025	Universidad Politécnica de Madrid
026	Universidad de Las Palmas de Gran Canaria
027	Universitat Politècnica de València
028	Universidad Nacional de Educación a Distancia
029	Universidad de Alcalá
030	Universidad de Deusto
031	Universidad de Navarra

032	Universidad Pontificia de Salamanca
033	Universidad Pontificia Comillas
034	Universidad de Castilla-La Mancha
035	Universidad Pública de Navarra
036	Universidad Carlos III de Madrid
037	Universidade da Coruña
038	Universidade de Vigo
039	Universitat Pompeu Fabra
040	Universitat Jaume I de Castelló
041	Universitat Ramón Llull
042	Universitat Rovira i Virgili
043	Universitat de Girona
044	Universitat de Lleida
045	Universidad de La Rioja
046	Universidad CEU San Pablo
047	Universidad Alfonso X El Sabio
048	Universidad de Almería
049	Universidad de Huelva
050	Universidad de Jaén
051	Universidad de Burgos
052	Universidad Antonio de Nebrija
053	Universidad Europea de Madrid
054	Universitat Oberta de Catalunya
055	Universidad Miguel Hernández
056	Universidad Rey Juan Carlos
057	Universidad IE
058	Universidad Pablo de Olavide
059	Universidad Católica Santa Teresa de Jesús de Ávila
060	Universitat de Vic- Universitat Central de Catalunya
061	Mondragon Unibertsitatea
062	Universitat Internacional de Catalunya
063	Universidad Internacional de Andalucía
064	Universidad Politécnica de Cartagena
065	Universidad Camilo José Cela
066	Universidad Católica San Antonio
067	Universidad CEU Cardenal Herrera

068	Universidad Francisco de Vitoria
069	Universidad Europea Miguel de Cervantes
070	Universitat Abat Oliba CEU
071	Universidad Internacional Menéndez Pelayo
072	Universidad Católica de Valencia San Vicente Mártir
073	Universidad San Jorge
074	Universidad a Distancia de Madrid
076	Valencian International University
077	Universidad Internacional de La Rioja
078	Universidad Tecnología y Empresa
079	Universidad Europea de Canarias
080	Universidad Internacional Isabel I de Castilla
081	Universidad Loyola Andalucía
082	Universidad Europea de Valencia

A.3. Código de SAS Base

A continuación, recogemos el código que hemos utilizado a lo largo del trabajo separado por los apartados según los objetivos.

1. Clústeres del perfil académico

```
/*Creamos la variable de idioma oficial*/
data trabajo.Datos;
set trabajo.Datos (drop = IDIOMA_OFICIAL);
if (TITIDI1 = 'No') & (TITIDI2='No') then IDIOMA_OFICIAL =
'No_TO_IDIOMA';
else IDIOMA_OFICIAL = 'Sí_TO_IDIOMA';
RUN;

/*Activos e inactivos;
MUESTREO ESTRATIFICADO ALEATORIO SIMPLE
SIN REEMPLAZAMIENTO*/
/*Población activa e inactiva*/
/*Cambiamos el tipo de la variable ID*/

DATA trabajo.todos (DROP=IDENT) ;
SET trabajo.datos;
IDENTI=put (IDENT,best10.);
run;

data trabajo.todos ;
set trabajo.todos ;
length EST_B1 $ 70; informat EST_B1 $ 70.; format EST_B1 $ 70.;
length EST2_MA $ 70; informat EST2_MA $ 70.; format EST2_MA $ 70.;
EST_B1="B_"||EST_B1;
EST2_MA="M_"||EST2_MA;
run;

/*DESDE AQUÍ*/
/*Estratificado*/
*MUESTREO ESTRATIFICADO ALEATORIO SIMPLE
SIN REEMPLAZAMIENTO;
data trabajo.todos (drop = IDENTI);
set trabajo.todos ;
IDEN = compress("O"||_n_);
run;

proc freq data=trabajo.todos;
tables TRBPRN1;run;

PROC SORT DATA=trabajo.todos
OUT=trabajo.todos1 ;
BY TRBPRN1 ;
RUN;
PROC SURVEYSELECT DATA=trabajo.todos1
OUT=trabajo.todos1tra N=20000 seed=12345 ;
STRATA TRBPRN1 / alloc=prop;
RUN;
proc freq data=trabajo.todos1tra;
tables TRBPRN1;run;
```

```

proc distance data=trabajo.todostra method=djaccard absent=0
out=trabajo.distjac;
var anominal(SEXO RAMA T_UNIV EST_B1 EST2_MA IDIOMA_OFICIAL);
id IDEN;
run;

proc cluster data=trabajo.distjac(type=distance) method=average pseudo
outtree=tree RSQUARE print=30;
var O4--O28904;
id IDEN;
run;

proc cluster data=trabajo.distjac(type=distance) method=ward pseudo
outtree=tree RSQUARE print=30;
var O4--O28904;
id IDEN;
run;

/*K-Modas en R, traemos el conjunto de datos*/
/*Datos cluster k modas R. 10 grupos escogidos*/

proc import datafile = 'A:\Aida\Tessy\Trabajo de Fin de
Mastercopia\clusterkmodas.xls'
out = trabajo.clusterkmodas
dbms = xls ;
run;

proc freq data=trabajo.clusterkmodas;
tables Cluster ;
run;

/*Caracterización
ACS*/

proc corresp data=trabajo.clusterkmodas all chi2p print=both DIMENS=3;
tables RAMA, Cluster;
run;

ods graphics on / width=500px height=330px border=off ;
proc corresp data=trabajo.union333 all chi2p print=both DIMENS=2;
tables idiomas, CLUSAC;
run;
proc corresp data=trabajo.union333 all chi2p print=both DIMENS=2;
tables TIC, CLUSAC;
run;

proc corresp data=trabajo.union333 all chi2p print=both DIMENS=2;
tables edad, CLUSAC;
run;

/*CARACTERIZACIÓN binarias*/
proc freq data=trabajo.union333;
tables EST_B1;
RUN;

PROC SORT DATA=trabajo.union333
OUT=trabajo.union333 ;
BY CLUSAC ;
RUN;

```

```

ods output onewayfreqs=trabajo.clacad_rep;
proc freq data=trabajo.union333;
tables sexo t_univ EST_B1
EST2_MA IDIOMA_OFICIAL;
by CLUSAC;
run;

proc transpose data= trabajo.clacad_rep out=trabajo.clacad_rep2;
by clusac;
var percent;
run;

/*Cambios hechos: missings y renombre variables*/
proc import datafile = 'A:\Aida\Tessy\Trabajo de Fin de
Mastercopia\cluscad_rep2.xls'
out = trabajo.cluscad_rep2
dbms = xls ;
run;

/*Heat Map prueba*/
proc sort data = trabajo.cluscad_rep2;
by CLUSAC;
run;

ods graphics on / width=820px height=590px border=off ;
data trabajo.heat_mapac(keep=VARIABLE CLUSTER ff);
array vector{6} ACAD_1 - ACAD_6;
set trabajo.cluscad_rep2;
a=0;
do aux='ACAD_1', 'ACAD_2', 'ACAD_3', 'ACAD_4', 'ACAD_5', 'ACAD_6' ;
a=a+1;
VARIABLE=CLUSAC;
CLUSTER=aux;
ff=vector{a};
output;
end;
run;

proc sgplot data=trabajo.heat_mapac;
heatmap x=VARIABLE y=CLUSTER/freq=ff colormodel=(cx42684e cxffffff
cx221c72);
Title "Porcentaje de la categoría en el clúster";
run;

```

2. Clústeres del perfil laboral

```

/*Este método en SAS nos está generando problemas, por lo que lo haremos
en R e importaremos el archivo*/
data trabajo.datos1;
set trabajo.datos1;
if TR_ANIO = '2015' then TR_ANIO1 = '2014 o más';
else if TR_ANIO = '2014' then TR_ANIO1 = '2014 o más';
else if TR_ANIO = '2013' then TR_ANIO1 = '2013';
else if TR_ANIO = '2012' then TR_ANIO1 = '2012';
else if TR_ANIO = '2011' then TR_ANIO1 = '2011';
else if TR_ANIO = 'No aplicable' then TR_ANIO1 = 'No aplicable';

```

```

else TR_ANIO = '2010 o menos';
run;

data trabajo.datos1;
set trabajo.datos1;
if BQ_NUM = '1' then BQ_NUM = '1';
else if BQ_NUM = '2' then BQ_NUM = '2';
else if BQ_NUM = '3' then BQ_NUM = '3';
else if BQ_NUM = '00' then BQ_NUM = 'NS/NC';
else if BQ_NUM = 'NS/NS' then BQ_NUM = 'Noaplicable';
else BQ_NUM = '4 o más';
RUN;

proc import datafile = 'A:\Aida\Tessy\Trabajo de Fin de
Master\datos1.xls'
  out = trabajo.datos1
  dbms = xls ;
run;

data trabajo.datos1;
set trabajo.datos1;
if LTR_PAI = 'España' then LTR_PAI = 'España';
else if LTR_PAI = 'Francia' then LTR_PAI = 'Otro país europeo';
else if LTR_PAI = 'Alemania' then LTR_PAI = 'Otro país europeo';
else if LTR_PAI = 'Otro país de la Unión Europea' then LTR_PAI = 'Otro
país europeo';
else if LTR_PAI = 'Otro país europeo (fuera de la UE)' then LTR_PAI =
'Otro país europeo';
else if LTR_PAI = 'Reino Unido' then LTR_PAI = 'Otro país europeo';
else if LTR_PAI = 'No aplicable' then LTR_PAI = 'No aplicable';
else LTR_PAI = 'Fuera de Europa';
RUN;

data trabajo.datos1;
length TRAB_F1 $ 70; informat TRAB_F1 $ 70.; format TRAB_F1 $ 70.;
set trabajo.datos1;
TRAB_F1="TF_"||TRAB_F1;
run;

proc freq data=trabajo.datos1;
tables BQ_NUM; run;

/*A partir de aquí datos1 está con las categorías agrupadas/

/*Clúster Laboral*/

/*Trabajando*/
data trabajo.trabajando;
set trabajo.datos1;
if TRBPRN1='Trabajando' then output trabajo.trabajando;
run;

/*Estratificado*/
proc freq data=trabajo.trabajando;
tables LTR_PAI;run;

PROC SORT DATA=trabajo.trabajando
  OUT=trabajo.trabajando1 ;
BY SIT_PRO ;
RUN;

```



```

PROC SURVEYSELECT DATA=trabajo.trabajando1
  OUT=trabajo.trabajandostra N=20000 seed=12345 ;
  STRATA SIT_PRO / alloc=prop;
RUN;
proc freq data=trabajo.trabajandostra;
tables SIT_PRO;run;
/*Matriz de distancias*/
proc distance data=trabajo.trabajandostra method=djaccard absent=0
out=trabajo.distjaclab;
var anominal(SEXO SIT_PRO JORNADA TR_CNAE BCOTC14 );
id IDEN;
run;

proc cluster data=trabajo.distjaclab(type=distance) method=ward pseudo
outtree=treet RSQUARE print=30;
var 011--028893;
id IDEN;
run;
proc cluster data=trabajo.distjaclab(type=distance) method=average
pseudo outtree=treesa RSQUARE print=30;
var 011--028893;
id IDEN;
run;

/*Pruebas de variables*/
data trabajo.trabajando2;
set trabajo.datos1;
if TRBPRN1='Trabajando' then output trabajo.trabajando2;
run;
data trabajo.trabajando22 (keep= IDEN SEXO SIT_PRO JORNADA TR_CNAE
BCOTC14 LTR_PA1 TRABOC TR_TAM TR_ANIO );
set trabajo.trabajando3;
run;
/*Estratificado*/
proc freq data=trabajo.trabajando2;
tables LTR_PA1;run;

PROC SORT DATA=trabajo.trabajando22
  OUT=trabajo.trabajando122 ;
BY SIT_PRO ;
RUN;
PROC SURVEYSELECT DATA=trabajo.trabajando122
  OUT=trabajo.trabajandostra22 N=20000 seed=12345 ;
  STRATA SIT_PRO / alloc=prop;
RUN;

/*Matriz de distancias*/
/*Cambiamos variables*/
proc distance data=trabajo.trabajandostra22 method=djaccard absent=0
out=trabajo.distjaclab22;
var anominal(SEXO SIT_PRO JORNADA TR_CNAE BCOTC14 TR_ANIO);id IDEN;
run;

proc cluster data=trabajo.distjaclab22(type=distance) method=ward
pseudo outtree=treet RSQUARE print=30;
var 011--028893;
id IDEN;
run;
/*Fin prueba de variables*/

/*Llevamos el archivo trabajando a R y realizamos el cluster k-modas.*/

```

```

/*Importamos el archivo de R tras el K-MODAS CLUSTER*/
proc import datafile = 'A:\Aida\Tessy\Trabajo de Fin de
Mastercopia\clusterkmodas_trab.xls'
out = trabajo.ckmodastrab
dbms = xls ;
run;
/*Cambiamos nombre cluster. 9 al 21*/
data trabajo.ckmodastrab;
set trabajo.ckmodastrab;
if CLUSTER = 1 then CLUSTER = 7;
else if CLUSTER = 2 then CLUSTER = 8;
else if CLUSTER = 3 then CLUSTER = 9;
else if CLUSTER = 4 then CLUSTER = 10;
else if CLUSTER = 5 then CLUSTER = 11;
else if CLUSTER = 6 then CLUSTER = 12;
else if CLUSTER = 7 then CLUSTER = 13;
else if CLUSTER = 8 then CLUSTER = 14;
else CLUSTER = 15;
run;

proc freq data=trabajo.ckmodastrab;
tables cluster;
run;

data trabajo.CT;
set trabajo.ckmodastrab;
run;
proc freq data=trabajo.ckmodastrab;
tables sit_pro;
run;

/*tiempo que ha estado trabajando, cuando encontró su primer empleo*/
/*titulados en junio 2010, encuesta en septiembre 2015, tiempo
trabajando*/

/*Desempleados*/
data trabajo.desempleados;
set trabajo.datos1;
if TRBPRN1='Desempleado' then output trabajo.desempleados;
run;
/*Desempleados no han trabajado*/
data trabajo.desempleadosno;
set trabajo.desempleados;
if HA_TRAB='No' then output trabajo.desempleadosno;
run;
/*Añadimos la variable cluster*/
data trabajo.desempleadosno;
set trabajo.desempleadosno;
CLUSTER = 2;
RUN;

proc freq data = trabajo.desempleados;
tables ;run;
/*Desempleados sí han trabajado*/
data trabajo.desempleadosi;
set trabajo.desempleados;
if HA_TRAB='Sí' then output trabajo.desempleadosi;
run;

/*Prueba variables para el clúster*/

```

```

proc distance data=trabajo.desempleadosi method=djaccard absent=0
out=trabajo.distjacdsi333;
var anominal(SEXO BQ_NUM TRAB_F1 );
id IDEN;
run;

proc cluster data=trabajo.distjacdsi333(type=distance) method=ward
pseudo outtree=treed RSQUARE print=15;
var O4--O28905;
id IDEN;
run;
/*Fin de prueba*/

/*Matriz de distancias*/
proc distance data=trabajo.desempleadosi method=djaccard absent=0
out=trabajo.distjacdsi;
var anominal(SEXO BQ_NUM BQ_TP_T TRAB_F1);
id IDEN;
run;
proc cluster data=trabajo.distjacdsi(type=distance) method=average
pseudo outtree=treedd RSQUARE print=15;
var O4--O28905;
id IDEN;
run;
proc cluster data=trabajo.distjacdsi(type=distance) method=ward pseudo
outtree=treed RSQUARE print=15;
var O4--O28905;
id IDEN;
run;

/*Llevamos el archivo a R para realizar el kmodas algoritmo*/
/*K-MODAS CLUSTER*/
proc import datafile = 'A:\Aida\Tessy\Trabajo de Fin de
Mastercopia\clusterkmodas_desem.xls'
out = trabajo.ckmodasdesem
dbms = xls ;
run;

/*Cambiamos nombre cluster. 3 al 8*/
data trabajo.ckmodasdesem;
set trabajo.ckmodasdesem;
if CLUSTER = 1 then CLUSTER = 3;
else if CLUSTER = 2 then CLUSTER = 4;
else if CLUSTER = 3 then CLUSTER = 5;
else CLUSTER = 6;
run;

data trabajo.CD;
set trabajo.ckmodasdesem;
run;

/*Inactivos*/
data trabajo.inactivos;
set trabajo.datos1;
if TRBPRN1='Inactivo' then output trabajo.inactivos;
run;
/*Añadimos la variable cluster*/
data trabajo.inactivos;
set trabajo.inactivos;
CLUSTER = 1;
RUN;

```

```

/*Unimos los tres archivos*/

data trabajo.CD;
set trabajo.CD;
if IDIOMAS = 'NA' then IDIOMAS = '';
IDIOMAS2 = IDIOMAS*1;
run;

data trabajo.CT;
set trabajo.CT;
if IDIOMAS = 'NA' then IDIOMAS = '';
IDIOMAS2 = IDIOMAS*1;
run;

data trabajo.CD (rename = (IDIOMAS2 = IDIOMAS)) ;
set trabajo.CD (drop = IDIOMAS);
RUN;

data trabajo.CT (rename = (IDIOMAS2 = IDIOMAS)) ;
set trabajo.CT (drop = IDIOMAS);
RUN;

data trabajo.claboral;
set trabajo.inactivos trabajo.CD trabajo.CT trabajo.desempleadosno;
RUN;

data trabajo.claboral (drop = BCOTC14_1) ;
set trabajo.claboral ;
RUN;

data trabajo.claboral ;
set trabajo.claboral ;
if substr(BCOTC14,1,4) = 'Cuar' then BCOTC14 = 'Cuarto quintil';
if substr(BCOTC14,1,4) = 'No a' then BCOTC14 = 'No aplicable';
if substr(BCOTC14,1,4) = 'Noap' then BCOTC14 = 'No aplicable';
if substr(BCOTC14,1,4) = 'Prim' then BCOTC14 = 'Primer quintil';
if substr(BCOTC14,1,4) = 'Quin' then BCOTC14 = 'Quinto quintil';
if substr(BCOTC14,1,4) = 'Terc' then BCOTC14 = 'Tercer quintil';
if substr(BCOTC14,1,4) = 'Segu' then BCOTC14 = 'Segundo quintil';
RUN;

proc freq data=trabajo.claboral;
tables sit_pro;
run;
/*R para arreglar los datos*/

proc import datafile = 'A:\Aida\Tessy\Trabajo de Fin de
Mastercopia\claboral.xls'
out = trabajo.claboral
dbms = xls ;
run;

data trabajo.claboral;
set trabajo.claboral;
if IDIOMAS = 'NA' then IDIOMAS = '';
IDIOMAS2 = IDIOMAS*1;
run;

```

```

data trabajo.claboral (rename = (IDIOMAS2 = IDIOMAS)) ;
set trabajo.claboral (drop = IDIOMAS);
RUN;

/*Caracterización*/
/*Trabajando*/

proc freq data=trabajo.ct;
tables SIT_PRO;
run;

proc corresp data=trabajo.ct all chi2p print=both DIMENS=3;
tables SIT_PRO, cluster;
run;

/*Heat map trabajadores, sexo y jornada*/
PROC SORT DATA=trabajo.ct
  OUT=trabajo.ct ;
BY cluster ;
RUN;

ods output onewayfreqs=trabajo.ct_rep;
proc freq data=trabajo.ct;
tables sexo JORNADA;
by cluster;
run;

proc freq data=trabajo.ct;
tables cluster;
RUN;

proc transpose data= trabajo.ct_rep out=trabajo.ct_rep2;
by cluster;
var percent;
run;

/*Cambios hechos: missings y renombre variables*/
proc import datafile = 'A:\Aida\Tessy\Trabajo de Fin de
Mastercopia\ct_rep2.xlsx'
  out = trabajo.ct_rep2
  dbms = xlsx ;
run;

/*grupos de variables, categorías frente a clúster*/

ods graphics on / width=820px height=590px border=off ;
data trabajo.heat_maptrab(keep=VARIABLE CLUSTER ff);
array vector{9} TRAB_7 - TRAB_15;
  set trabajo.ct_rep2;
  a=0;
  do aux='TRAB_ 7','TRAB_ 8','TRAB_ 9','TRAB_10', 'TRAB_11', 'TRAB_12',
'TRAB_13' , 'TRAB_14', 'TRAB_15';
    a=a+1;
    VARIABLE=CLUSTER_T;
    CLUSTER=aux;
    ff=vector{a};
    output;
  end;
run;

```

```

proc sgplot data=trabajo.heat_maptrab;
  heatmap x=VARIABLE y=CLUSTER/freq=ff colormodel=(cx42684e cxffffff
cx221c72);
  Title "Porcentaje de la categoría en el clúster";
run;

/*CARACTERIZACIÓN DE LOS DESEMPLEADOS*/

/*Si han trabajado*/
proc freq data = trabajo.cd;
tables cluster;
run;
proc corresp data=trabajo.cd all chi2p print=both DIMENS=3;
tables BQ_NUM, cluster;
run;
/*Cambiamos el nombre a las categorías para facilitad la interpretación*/
data trabajo.cd1 ;
set trabajo.cd ;
run;
data trabajo.cd1 ;
set trabajo.cd1 ;
if BQ_NUM = '1' then BQ_NUM = 'Uno';
if BQ_NUM = '2' then BQ_NUM = 'Dos';
if BQ_NUM = '3' then BQ_NUM = 'Tres';
else BQ_NUM = BQ_NUM;
RUN;
proc freq data = trabajo.cd1;
tables BQ_NUM;
run;

proc corresp data=trabajo.cd1 all chi2p print=both DIMENS=3;
tables BQ_NUM, cluster;
run;

proc corresp data=trabajo.cd all chi2p print=both DIMENS=2;
tables edad, cluster;
run;

/*Heat-map
TRAB_F1 SEXO*/

PROC SORT DATA=trabajo.cd
  OUT=trabajo.cd ;
BY cluster ;
RUN;

ods output onewayfreqs=trabajo.cd_rep;
proc freq data=trabajo.cd;
tables TRAB_F1 SEXO;
by cluster;
run;

proc transpose data= trabajo.cd_rep out=trabajo.cd_rep2;
by cluster;
var percent;
run;

/*Cambios hechos: missings y renombre variables*/
proc import datafile = 'A:\Aida\Tessy\Trabajo de Fin de
Mastercopia\cd_rep2.xlsx'

```

```

out = trabajo.cd_rep2
dbms = xlsx ;
run;

/*grupos de variables, categorías frente a clúster*/
PROC SORT DATA=trabajo.cd_rep2
OUT=trabajo.cd_rep2 ;
BY CLUSTER_D ;
RUN;

ods graphics on / width=820px height=590px border=off ;
data trabajo.heat_mapdes(keep=VARIABLE CLUSTER ff);
array vector{4} DESEM_SI_3 - DESEM_SI_6;
set trabajo.cd_rep2;
a=0;
do aux='DESEM_SI_3', 'DESEM_SI_4', 'DESEM_SI_5', 'DESEM_SI_6' ;
a=a+1;
VARIABLE=CLUSTER_D;
CLUSTER=aux;
ff=vector{a};
output;
end;
run;

proc sgplot data=trabajo.heat_mapdes;
heatmap x=VARIABLE y=CLUSTER/freq=ff colormodel=(cx42684e cxffffff
cx221c72);
Title "Porcentaje de la categoría en el clúster";
run;

/*DESEMPLEADOS QUE NO HAN TRABAJADO NUNCA*/
proc freq data=trabajo.desempleadosno;
tables BSQDIF3;
run;

proc sgplot data=trabajo.heat_mapdes;
heatmap x=VARIABLE y=CLUSTER/freq=ff colormodel=(cx42684e cxffffff
cx221c72);
Title "Porcentaje de la categoría en el clúster";
run;

ods graphics on / width=500px height=290px border=off ;
proc gchart data=trabajo.desempleadosno1;
vbar BSQDIF3 / type=percent;
Title "Preparando oposiciones";
run;

data trabajo.desempleadosno1 ;
set trabajo.desempleadosno ;
IF BSQC025 in ('NS/NC') THEN delete;
run;

proc freq data=trabajo.desempleadosno1;
tables BSQDIF3;
run;

/*INACTIVOS*/
proc freq data=trabajo.inactivos;
tables BSQCOM8
;
run;

```

3. Análisis de Correspondencias Simples

```
/*En cuanto a la variable sexo está todo*/
/*Ya están las variables que han sido reagrupadas: TR_ANIO, LPR_PAI,
BQ_NUM en union333*/

data trabajo.CLABORAL2(rename = (cluster = clusterlab));
set trabajo.CLABORAL;
run;

proc freq data=trabajo.claboral2;
tables CLUSLAB;
run;

data trabajo.claboral2;
length clusterlab $ 70;informat clusterlab $ 70.; format clusterlab $
70.;
set trabajo.claboral2;
run;

data trabajo.claboral2;
set trabajo.claboral2;
CLUSLAB = compress("LAB_"||clusterlab);
run;

data trabajo.cacademico;
set trabajo.clusterkmodas;
CLUSAC = compress("ACAD_"||cluster);
run;

proc freq data=trabajo.cacademico;
tables CLUSAC;
run;

proc sort data = trabajo.CLABORAL2;
by iden;
run;
proc sort data = trabajo.cacademico;
by iden;
run;

data trabajo.union333 (drop= clusterlab CLUSTER);
merge trabajo.CLABORAL2 trabajo.cacademico;
by iden;
run;

data trabajo.union (keep = iden CLUSLAB CLUSAC);
merge trabajo.CLABORAL2 trabajo.cacademico;
by iden;
run;

proc freq data=trabajo.union333;
tables CLUSLAC;
run;
```



```

proc corresp data=trabajo.union333 print=both DIMENS=2;
tables CLUSLAB, CLUSAC;
run;

proc freq data=trabajo.union333;
tables TR_ANIO;
run;

/*Análisis de correspondencias*/
data trabajo.union (keep = IDEN CLUSLAB CLUSAC);
set trabajo.union333;
run;

proc freq data=trabajo.union;
tables CLUSLAB CLUSAC;
run;

proc corresp data=trabajo.union all print=both chi2p DIMENS=2;
ods output CellChiSq = Aportaciones;
tables CLUSLAB, CLUSAC;
run;

data Aportaciones2(drop=Sum);
set Aportaciones;
if Label="Sum" then delete;
run;
data Aportaciones3(keep=filas col ff);
array vector{10} ACAD_1 ACAD_2 ACAD_3 ACAD_4 ACAD_5 ACAD_6 ACAD_7 ACAD_8
ACAD_9 ACAD_10 ;
set Aportaciones2;
a=0;
do aux='ACAD_1', 'ACAD_2', 'ACAD_3', 'ACAD_4', 'ACAD_5', 'ACAD_6',
'ACAD_7', 'ACAD_8', 'ACAD_9', 'ACAD_10' ;
a=a+1;
filas=label;
col=aux;
ff=vector{a};
output;
end;
run;
proc sgplot data=Aportaciones3;
heatmap x=filas y=col/freq=ff colormodel=(cxxxxxxx cx42684e cx221c72
);
Title "Aportaciones a chi^2";
run;

proc sgplot data=inercia;
series x=ID y=Inertia;
where ID<=3;
xaxis label="Numero factores";
run;

```

A.4. Código de R

A continuación, recogemos el código que hemos utilizado a lo largo del trabajo en RStudio para el desarrollo del Análisis Clúster no Jerárquico y parte de la depuración del conjunto de datos.

- Función *k-modes*

```
kmodes <- function(data, modes, iter.max = 10, weighted =
FALSE, fast = TRUE) {
  if(!is.data.frame(data)) data <- as.data.frame(data)
  # check data for coloumn types
  isnumeric <- sapply(data, is.numeric)
  isfactor <- sapply(data, is.factor)
  if(any(isfactor)){
    levs <- vector("list", ncol(data))
    for(j in which(isfactor)){
      levsj <- levels(data[,j])
      data[,j] <- levsj[data[,j]] # replace factors by
characters (for computation of distances if weighted == TRUE)
      levs[[j]] <- levsj # store levels for back transform
of modes to factor (otherwise some levels might get lost)
    }
  }

  # check for numerics where the values potentially do not
represent categories
  if(any(isnumeric)) {
    lengths <- sapply(data[,isnumeric], function(z)
return(length(unique(z))))
    if(any(lengths > 30)) warning("data has numeric coloumns
with more than 30 different levels!")
  }

  ### updates the mode of cluster "num",
  ### is called everytime an object switches clusters
```

```

update_mode <- function(num, num_var, data, cluster) {
  ## gather all objects of cluster "num"
  clust <- data[which(cluster == num),]
  apply(clust, 2, function(cat) {
    ## compute the most frequent category for each variable
    cat <- table(cat)
    names(cat)[which.max(cat)]
  })
} # end update mode

### computes the weighted distance between an object "obj"
and the mode of a cluster

### the frequencies of categories in "data" are used for
the weighting

distance <- function(mode, obj, weights){
  if(is.null(weights))
    return(sum(mode != obj))

  #obj <- as.integer(obj) # may lead to wrong calc. of
ln.51/52 (no match)
  obj <- as.character(obj)
  mode <- as.character(mode)
  different <- which(mode != obj)
  n_mode <- n_obj <- numeric(length(different))
  for (i in seq(along = different)) {
    ## frequencies are computed only if neccessary (not if
distance is 0 anyway)
    weight <- weights[[different[i]]]
    names <- names(weight)
    n_mode[i] <- weight[which(names==mode[different[i]])]
    n_obj[i] <- weight[which(names==obj[different[i]])]
  }
}

```

```

    }
    dist <- sum((n_mode + n_obj) / (n_mode * n_obj))
    return(dist)
} # end distance

n <- nrow(data)
num_var <- ncol(data)
data <- as.data.frame(data)

cluster <- numeric(n) # the actual partition
names(cluster) <- 1:n

if (missing(modes))
  stop("'modes' must be a number or a data frame.")
if (iter.max < 1)
  stop("'iter.max' must be positive.")

if (length(modes) == 1) {
  ## if the desired number "k" of clusters is given, choose
  k object randomly from "data"
  k <- modes
  modes <- unique(data)[sample(nrow(unique(data)))[1:k],]
  for (i in 1:k)
    cluster[which(rownames(data) == rownames(modes)[i])]
<- i
} else {
  ## use given modes if appropriate
  if (any(duplicated(modes)))
    stop("Initial modes are not distinct.")
  if (ncol(data) != ncol(modes))

```

```

    stop("'data' and 'modes' must have same number of
columns")

    #modes <- as.matrix(modes) # replaced

    modes <- as.data.frame(modes)

    if(any(isfactor)){ # here further checks on could be
added...

        if(!all(sapply(modes[,isfactor],          is.factor)))
stop("Types of modes do not match data!")

        for(j      in      which(isfactor))      modes[,j]      <-
levels(modes[,j])[modes[,j]]

    }

    for(j in 1:num_var) if(weighted) if(!all(modes[,j] %in%
unique(data[,j]))) stop("For weighted call values of modes
must exist in data!")

    k <- nrow(modes)

    }

    if (k > nrow(unique(data)))

        stop("More cluster modes than distinct data points.")

    if(weighted){

        ## compute the frequencies of each category for each
variable

        weights <- vector("list", num_var)

        for (i in 1:num_var)

            weights[[i]] <- table(data[,i])

    } else {

        weights <- NULL

    }

    # original implementation

    if(!fast){

        for (j in which(cluster==0)) {

            ## first put all not yet assigned objects into the
cluster, which has the nearest mode

```

```

        dist <- apply(modes, 1, function(x) distance(x,
data[j,], weights))

        cluster[j] <- which.min(dist)

        modes[cluster[j],] <- update_mode(cluster[j],
num_var, data, cluster)

        ## update the mode of the cluster the object has been
assigned to
    }

    for (i in 1:iter.max) {
        continue <- FALSE
        for (j in 1:n) {
            ## run through all objects and assign them to the
cluster with the nearest mode

            ## (or leave everything as is, if the object's
current cluster's mode is still the nearest)

            dist <- apply(modes, 1, function(x) distance(x,
data[j,], weights))

            clust_new <- which.min(dist)

            clust_old <- cluster[j]

            if (clust_new != clust_old) {
                ## update the modes of old and new cluster, if
object has switched from "clust_old" to "clust_new"

                cluster[j] <- clust_new

                modes[clust_new,] <- update_mode(clust_new,
num_var, data, cluster)

                modes[clust_old,] <- update_mode(clust_old,
num_var, data, cluster)

                continue <- TRUE
            }
        }
    }

    ## break, if during a complete iteration no object has
changed clusters

    if (!continue) break
} # end iterations

```

```

}

### fast version of the algorithm

### ...running through the entire data set at once before
updating clusters which allows for apply
if(fast){

  ## first put all not yet assigned objects into the
  cluster, which has the nearest mode

  # compute distances: either weighted or unweighted
  dists <- matrix(NA, nrow=n, ncol = k)
  if(!weighted){
    for(i in 1:k){
      di      <-      sapply(1:ncol(data),      function(j)
return(data[,j] != rep(modes[i,j], n)) )
      di <- rowSums(di)
      dists[,i] <- di
    }
  }
  if(weighted){

    # compute frequencies (for weighted computation only
    once necessary)

    n_obj <- matrix(NA, nrow=n, ncol = ncol(data))

    for(j in 1:ncol(data)) n_obj[,j] <-
weights[[j]][sapply(as.character(data[,j]), function(z)
return(which(names(weights[[j]])==z))))]

    # necessary for updated modes after each iteration

    n_mode <- matrix(NA, nrow=nrow(modes), ncol =
ncol(data))

    for(j in 1:ncol(data)) n_mode[,j] <-
weights[[j]][sapply(as.character(modes[,j]), function(z)
return(which(names(weights[[j]])==z))))]

    for(i in 1:k){

```

```

        di      <-      sapply(1:ncol(data),      function(j)
return(data[,j] != rep(modes[i,j], n)) )

        wts      <-      (n_mode[rep(i,n),]      +      n_obj)      /
(n_mode[rep(i,n),] * n_obj)

        di <- rowSums(di*wts)

        dists[,i] <- di

    }

}

# assign clusters

#old.cluster <- cluster

cluster      <-      apply(dists, 1, function(z) {a <-
which.min(z); if (length(a)>1) a <- sample(a,1); return(a)})
# sample in case of multiple minima

#min.dists      <-      apply(cbind(clusters, dists), 1,
function(z) z[z[1]+1])

# update modes

for(j in 1:nrow(modes)) modes[j,] <- update_mode(j,
num_var, data, cluster)


# start iterations
for (i in 1:iter.max) {
    continue <- FALSE

    # compute distances: either weighted or unweighted

    dists <- matrix(NA, nrow=n, ncol = k)

    if(!weighted){
        for(i in 1:k){
            di      <-      sapply(1:ncol(data),      function(j)
return(data[,j] != rep(modes[i,j], n)) )

            di <- rowSums(di)

            dists[,i] <- di

        }

    }

    if(weighted){

```



```

# necessary for updated modes after each iteration

n_mode <- matrix(NA, nrow=nrow(modes), ncol =
ncol(data))

for(j in 1:ncol(data)) n_mode[,j] <-
weights[[j]][sapply(as.character(modes[,j]), function(z)
return(which(names(weights[[j]])==z)))]]

for(i in 1:k){
  di <- sapply(1:ncol(data), function(j)
return(data[,j] != rep(modes[i,j], n)) )
  wts <- (n_mode[rep(i,n),] + n_obj) /
(n_mode[rep(i,n),] * n_obj)
  di <- rowSums(di*wts)
  dists[,i] <- di
}
}

# assign clusters

old.cluster <- cluster

cluster <- apply(dists, 1, function(z) {a <-
which.min(z); if (length(a)>1) a <- sample(a,1); return(a)})
# sample in case of multiple minima

#min.dists <- apply(cbind(clusters, dists), 1,
function(z) z[z[1]+1])

# update modes

for(j in 1:nrow(modes)) modes[j,] <- update_mode(j,
num_var, data, cluster)

## break, if during a complete iteration no object has
changed clusters

if(any(old.cluster != cluster)) continue <- TRUE

if (!continue) break

} # end iterations

}

# compute results

```

```

cluster.size <- table(cluster) # sizes of all clusters
if (length(cluster.size) < k)
  warning("One or more clusters are empty.")

# compute final matrix of distances: either weighted or
unweighted

dists <- matrix(NA, nrow=n, ncol = k)

if(weighted){
  n_mode <- matrix(NA, nrow=nrow(modes), ncol =
ncol(data))

  for(j in 1:ncol(data)) n_mode[,j] <-
weights[[j]][sapply(as.character(modes[,j]), function(z)
return(which(names(weights[[j]])==z))))]
}

for(i in 1:k){
  di <- sapply(1:ncol(data), function(j) return(data[,j]
!= rep(modes[i,j], n)) )

  if(!weighted) di <- rowSums(di)

  if(weighted){
    if(!fast){# ...for fast == FALSE n_obj ist computed
within calls of distance()

      n_obj <- matrix(NA, nrow=n, ncol = ncol(data))

      for(j in 1:ncol(data)) n_obj[,j] <-
weights[[j]][sapply(as.character(data[,j]), function(z)
return(which(names(weights[[j]])==z))))]
    }

    wts <- (n_mode[rep(i,n),] + n_obj) /
(n_mode[rep(i,n),] * n_obj)

    di <- rowSums(di*wts)
  }

  dists[,i] <- di
}

```

```

    ## compute for each cluster the sum of the distances of
    all objects of a cluster to the cluster's mode:

    diffs <- numeric(k)

    for (i in seq_along(cluster.size)) diffs[i] <-
sum(dists[cluster == i,i])

    #diffs[i] <- sum(apply(data[cluster == i,], 1, function(x)
sum(x != modes[i,]) ))

    rownames(modes) <- 1:k
    colnames(modes) <- colnames(data)

    # convert coloumn modes to original coloumn types of data
    if(any(isfactor)) for(j in which(isfactor)) modes[,j] <-
factor(modes[,j], levels = levs[[j]])

    if(any(isnumeric)) for(j in which(isnumeric)) modes[,j] <-
as.numeric(modes[,j])

    result <- list(cluster = cluster, size = cluster.size,
modes = modes,

                    withindiff = diffs, iterations = i,
weighted = weighted)

    class(result) <- "kmodes"

    return(result)
}

print.kmodes <- function(x, ...)
{
    cat("K-modes clustering with ", length(x$size), " clusters
of sizes ",

        paste(x$size, collapse=", "), "\n", sep="")
    cat("\nCluster modes:\n")
    print(x$modes, ...)
    cat("\nClustering vector:\n")

```

```

print(x$cluster, ...)
cat("\nWithin cluster simple-matching distance by
cluster:\n")

print(x$withindiff, ...)
cat("\nAvailable components:\n")
print(names(x))
invisible(x)
}

```

- Clúster académico

```

library(sas7bdat)

datos <- read.sas7bdat('A:\\Aida\\Tessy\\Trabajo de Fin de
Mastercopia\\Datos1.sas7bdat')

todos <- subset(datos, select = c(SEXO, TRBPRN1, RAMA,
T_UNIV, EST_B1, EST2_MA, IDIOMA_OFICIAL))

todos2 <- subset(datos, select = c(IDEN, SEXO, TRBPRN1,
RAMA, T_UNIV, EST_B1, EST2_MA, IDIOMA_OFICIAL))

cluster <- kmodes(todos, 6, iter.max = 400, weighted = FALSE,
fast = TRUE)

nueval <- cbind(todos2, CLUSTER=cluster$cluster)
for (i in 1:nrow(nueval)){
if (nueval$T_UNIV[i] == 'Universidad\r\nPrivada'){
nueval$T_UNIV2[i] <- 'Universidad Pública'
} else {
nueval$T_UNIV2[i] <- 'Universidad Privada'
}
}

for (i in 1:nrow(nueval)){
if (nueval$RAMA[i] == 'Artes y\r\nhumanidades'){
nueval$RAMA2[i] <- 'Artes y humanidades'
} else if(nueval$RAMA[i] == 'Ciencias de la\r\nsalud') {
nueval$RAMA2[i] <- 'Ciencias de la salud'
}
}

```

```

    }else if(nueval$RAMA[i] == 'Ciencias\r\nsociales
y\r\njurídicas') {
        nueval$RAMA2[i] <-'Ciencias sociales y jurídicas'
    }else if(nueval$RAMA[i] == 'Ingeniería y\r\narquitectura')
    {
        nueval$RAMA2[i] <-'Ingeniería y arquitectura'
    }else {
        nueval$RAMA2[i] <-'Ciencias'
    }
}

```

```

Write.csv2(nueval, 'A:\\Aida\\Tessy\\Trabajo de Fin de
Mastercopia\\clusterkmodas.csv')

```

```

install.packages("readxl")

```

```

library(readxl)

```

```

clusterkmodas <- read_excel("A:\\Aida\\Tessy\\Trabajo de Fin
de Master\\clusterkmodas.xlsx")

```

- Clúster laboral

```

#TRABAJANDO

```

```

trabajando <- read.sas7bdat('A:\\Aida\\Tessy\\Trabajo de Fin
de Mastercopia\\trabajando.sas7bdat')

```

```

trab <- subset(trabajando, select = c(SEXO,SIT_PRO, JORNADA,
TR_CNAE, BCOTC14))

```

```

cluster_trab <- kmodes(trab, 9, iter.max = 400, weighted =
FALSE, fast = TRUE)

```

```

nueva_trab <- cbind(trabajando,
CLUSTER=cluster_trab$cluster)

```

```

for (i in 1:nrow(nueva_trab)){
    if (nueva_trab$JORNADA[i] == 'A tiempo\r\ncompleto'){

```

```

nueva_trab$JORNADA2[i] <- 'A tiempo completo'
} else {
nueva_trab$JORNADA2[i] <- 'A tiempo parcial'
}
}
for (i in 1:nrow(nueva_trab)){
  if (nueva_trab$SIT_PRO[i] == 'Asalariado con
contrato\r\npermanente'){
nueva_trab$SIT_PRO2[i] <- 'Asalariado con contrato
permanente'
} else if(nueva_trab$SIT_PRO[i] == 'Asalariado con
contrato\r\ntemporal') {
nueva_trab$SIT_PRO2[i] <- 'Asalariado con contrato
temporal'
} else if(nueva_trab$SIT_PRO[i] == 'Empresario o
trabajador\r\nindependiente') {
nueva_trab$SIT_PRO2[i] <- 'Empresario o trabajador
independiente'
} else if(nueva_trab$SIT_PRO[i] == 'En prácticas,
formación\r\nno becario') {
nueva_trab$SIT_PRO2[i] <- 'En prácticas, formación o
becario'
} else {
nueva_trab$SIT_PRO2[i] <- 'Ayuda familiar'
}
}

for (i in 1:nrow(nueva_trab)){
  if (nueva_trab$TR_CNAE[i] == 'Actividades administrativas
de oficina y\r\notras actividades auxiliares a
las\r\nempresas'){
nueva_trab$TR_CNAE2[i] <- 'Actividades administrativas
de oficina y otras actividades auxiliares a las empresas'

```

```

    } else if(nueva_trab$TR_CNAE[i] == 'Actividades
artísticas, culturales,\r\nrecreativas, deportivas y de
entrenimiento') {

        nueva_trab$TR_CNAE2[i] <-'Actividades artísticas,
culturales, recreativas, deportivas y de entrenamiento'

    } else if(nueva_trab$TR_CNAE[i] == 'Actividades de
agencias\r\nviajes, operadores
turísticos,servicios\r\nreservas y actividades
relacionadas') {

        nueva_trab$TR_CNAE2[i] <-'Actividades de agencias
viajes, operadores turísticos,servicios reservas y
actividades relacionadas'

    } else if(nueva_trab$TR_CNAE[i] == 'Actividades de
organizaciones y\r\norganismos extraterritoriales') {

        nueva_trab$TR_CNAE2[i] <-'Actividades de organizaciones
y organismos extraterritoriales'

    } else if(nueva_trab$TR_CNAE[i] == 'Actividades
profesionales, científicas y\r\ntécnicas') {

        nueva_trab$TR_CNAE2[i] <-'Actividades profesionales,
científicas y técnicas'

    } else if(nueva_trab$TR_CNAE[i] == 'Actividades sanitarias
y de servicios\r\nsociales') {

        nueva_trab$TR_CNAE2[i] <-'Actividades sanitarias y de
servicios sociales'

    } else if(nueva_trab$TR_CNAE[i] == 'Administración Pública
y defensa;\r\nSeguridad Social obligatoria') {

        nueva_trab$TR_CNAE2[i] <-'Administración Pública y
defensa;Seguridad Social obligatoria'

    } else if(nueva_trab$TR_CNAE[i] == 'Agricultura,
ganadería, silvicultura y\r\npesca') {

        nueva_trab$TR_CNAE2[i] <-'Agricultura, ganadería,
silvicultura y pesca'

    } else if(nueva_trab$TR_CNAE[i] == 'Comercio al por mayor
y al por menor;\r\nreparación de vehículos de motor
y\r\nmotocicletas') {

        nueva_trab$TR_CNAE2[i] <-'Comercio al por mayor y al por
menor; reparación de vehículos de motor y motocicletas'

```

```

    } else if(nueva_trab$TR_CNAE[i] == 'Efectos personales y
artículos de uso\r\ndoméstico y otros servicios personales')
{
    nueva_trab$TR_CNAE2[i] <-'Efectos personales y artículos
de uso doméstico y otros servicios personales'

    } else if(nueva_trab$TR_CNAE[i] == 'Particulares como
empleadores
de\r\npersonal\r\ndoméstico;particulares:productores
de\r\nbbss de uso propio') {

    nueva_trab$TR_CNAE2[i] <-'Particulares como empleadores
de personal doméstico;particulares:productores de bbss de
uso propio'

    } else if(nueva_trab$TR_CNAE[i] == 'Servicios a edificios
y actividades de\r\njardinería') {

    nueva_trab$TR_CNAE2[i] <-'Servicios a edificios y
actividades de jardinería'

    } else if(nueva_trab$TR_CNAE[i] == 'Suministro de agua,
actividades de\r\nsaneamiento, gestión de residuos
y\r\ndescontaminación') {

    nueva_trab$TR_CNAE2[i] <-'Suministro de agua,
actividades de saneamiento, gestión de residuos y
descontaminación'

    } else if(nueva_trab$TR_CNAE[i] == 'Suministro de energía
eléctrica, gas,\r\nvapor y aire acondicionado') {

    nueva_trab$TR_CNAE2[i] <-'Suministro de energía
eléctrica, gas, vapor y aire acondicionado'

    } else if(nueva_trab$TR_CNAE[i] == 'Transporte,
almacenamiento, actividades\r\npostales y de correos') {

    nueva_trab$TR_CNAE2[i] <-'Transporte, almacenamiento,
actividades postales y de correos'

    } else if(nueva_trab$TR_CNAE[i] == 'Actividades
asociativas y sindicales') {

    nueva_trab$TR_CNAE2[i] <-'Actividades asociativas y
sindicales'

    } else if(nueva_trab$TR_CNAE[i] == 'Actividades de
fotografía') {

    nueva_trab$TR_CNAE2[i] <-'Actividades de fotografía'

    } else if(nueva_trab$TR_CNAE[i] == 'Actividades de
mantenimiento físico') {

```



```

    nueva_trab$TR_CNAE2[i] <-'Actividades de mantenimiento
físico'

    } else if(nueva_trab$TR_CNAE[i] == 'Actividades de
seguridad e investigación') {

        nueva_trab$TR_CNAE2[i] <-'Actividades de seguridad e
investigación'

    } else if(nueva_trab$TR_CNAE[i] == 'Actividades
financieras y de seguros') {

        nueva_trab$TR_CNAE2[i] <-'Actividades financieras y de
seguros'

    } else if(nueva_trab$TR_CNAE[i] == 'Actividades
inmobiliarias') {

        nueva_trab$TR_CNAE2[i] <-'Actividades inmobiliarias'

    } else if(nueva_trab$TR_CNAE[i] == 'Actividades
relacionadas con el empleo') {

        nueva_trab$TR_CNAE2[i] <-'Actividades relacionadas con
el empleo'

    } else if(nueva_trab$TR_CNAE[i] == 'Actividades
veterinarias') {

        nueva_trab$TR_CNAE2[i] <-'Actividades veterinarias'

    } else if(nueva_trab$TR_CNAE[i] == 'Construcción e
ingeniería civil') {

        nueva_trab$TR_CNAE2[i] <-'Construcción e ingeniería
civil'

    } else if(nueva_trab$TR_CNAE[i] == 'Educación') {

        nueva_trab$TR_CNAE2[i] <-'Educación'

    } else if(nueva_trab$TR_CNAE[i] == 'Hostelería') {

        nueva_trab$TR_CNAE2[i] <-'Hostelería'

    } else if(nueva_trab$TR_CNAE[i] == 'Industria
manufacturera') {

        nueva_trab$TR_CNAE2[i] <-'Industria manufacturera'

    } else if(nueva_trab$TR_CNAE[i] == 'Industrias
extractivas') {

        nueva_trab$TR_CNAE2[i] <-'Industrias extractivas'

    } else if(nueva_trab$TR_CNAE[i] == 'Información y
comunicaciones') {

```

```

    nueva_trab$TR_CNAE2[i] <-'Información y comunicaciones'
  } else if(nueva_trab$TR_CNAE[i] == 'NS/NC') {
    nueva_trab$TR_CNAE2[i] <-'NS/NC'
  } else {
    nueva_trab$TR_CNAE2[i] <-'Reparación de ordenadores'
  }
}

for (i in 1:nrow(nueva_trab)){
  if (nueva_trab$BCOTC14[i] == 'Cuarto\r\nquintil'){
    nueva_trab$BCOTC142[i] <- 'Cuarto quintil'
  } else if(nueva_trab$BCOTC14[i] == 'No\r\naplicable') {
    nueva_trab$BCOTC142[i] <-'No aplicable'
  } else if(nueva_trab$BCOTC14[i] == 'Primer\r\nquintil') {
    nueva_trab$BCOTC142[i] <-'Primer quintil'
  } else if(nueva_trab$BCOTC14[i] == 'Quinto\r\nquintil') {
    nueva_trab$BCOTC142[i] <-'Quinto quintil'
  } else if(nueva_trab$BCOTC14[i] == 'Segundo\r\nquintil')
  {
    nueva_trab$BCOTC142[i] <-'Segundo quintil'
  } else if(nueva_trab$BCOTC14[i] == 'Tercer\r\nquintil') {
    nueva_trab$BCOTC142[i] <-'Tercer quintil'
  } else {
    nueva_trab$BCOTC142[i] <-'NS/NC'
  }
}

write.csv2(nueva_trab, 'A:\\Aida\\Tessy\\Trabajo de Fin de
Mastercopia\\clusterkmodas_trab.csv')

#DESEMPLEADOS

```

```

desemple <- read.sas7bdat('A:\\Aida\\Tessy\\Trabajo de Fin
de Mastercopia\\desempleadosi.sas7bdat')

desem <- subset(desemple, select = c(SEXO, BQ_NUM, BQ_TP_T
, TRAB_F1))

cluster_desem <- kmodes(desem, 4, iter.max = 400, weighted
= FALSE, fast = TRUE)

nueva_desem <- cbind(desemple,
CLUSTER=cluster_desem$cluster)

for (i in 1:nrow(nueva_desem)){
  if (nueva_desem$BQ_TP_T[i] == '2 o más\\r\\naños') {
    nueva_desem$BQ_TP_T2[i] <- '2 o más años'
  } else if(nueva_desem$BQ_TP_T[i] == 'De 1 año y\\r\\nmedio
a 2\\r\\naños') {
    nueva_desem$BQ_TP_T2[i] <- 'De 1 año y medio a 2 años'
  } else if(nueva_desem$BQ_TP_T[i] == 'De 1 año\\r\\na año
y\\r\\nmedio') {
    nueva_desem$BQ_TP_T2[i] <- 'De 1 año a año y medio'
  } else if(nueva_desem$BQ_TP_T[i] == 'De 6\\r\\nmeses a
1\\r\\naño') {
    nueva_desem$BQ_TP_T2[i] <- 'De 6 meses a 1 año'
  } else if(nueva_desem$BQ_TP_T[i] == 'Menos de\\r\\n6 meses')
{
    nueva_desem$BQ_TP_T2[i] <- 'Menos de 6 meses'
  } else {
    nueva_desem$BQ_TP_T2[i] <- 'NS/NC'
  }
}

write.csv2(nueva_desem, 'A:\\Aida\\Tessy\\Trabajo de Fin de
Mastercopia\\clusterkmodas_desem.csv')

```

#Importamos el archivo de sas con el total de los datos para corregir una agrupación que está generando error en SAS ya que no lo hace correctamente.

```
library(sas7bdat)
```

```
datos <- read.sas7bdat('A:\\Aida\\Tessy\\Trabajo de Fin de Master\\datos.sas7bdat')
```

```
for (i in 1:nrow(datos)){  
  if (datos$TR_ANIO[i] == '2015') {  
    datos$TR_ANIO1[i] <- '2014 o más'  
  } else if(datos$TR_ANIO[i] == '2014') {  
    datos$TR_ANIO1[i] <- '2014 o más'  
  } else if(datos$TR_ANIO[i] == '2013') {  
    datos$TR_ANIO1[i] <- '2013'  
  } else if(datos$TR_ANIO[i] == '2012') {  
    datos$TR_ANIO1[i] <- '2012'  
  } else if(datos$TR_ANIO[i] == '2011') {  
    datos$TR_ANIO1[i] <- '2011'  
  } else if(datos$TR_ANIO[i] == 'No\\naplicable') {  
    datos$TR_ANIO1[i] <- 'No aplicable'  
  } else {  
    datos$TR_ANIO1[i] <- '2010 o menos'  
  }  
}
```

```
for (i in 1:nrow(datos)){  
  if (datos$BQ_NUM[i] == '00') {  
    datos$BQ_NUM1[i] <- 'NS/NC'  
  } else if(datos$BQ_NUM[i] == '01') {  
    datos$BQ_NUM1[i] <- '01'  
  } else if(datos$BQ_NUM[i] == '02') {  
    datos$BQ_NUM1[i] <- '02'  
  } else if(datos$BQ_NUM[i] == '03') {
```

```

    datos$BQ_NUM1[i] <-'03'
  } else if(datos$BQ_NUM[i] == 'No\naplicable') {
    datos$BQ_NUM1[i] <-'No aplicable'
  } else if(datos$BQ_NUM[i] == '00') {
    datos$BQ_NUM1[i] <-'NS/NC'
  } else if(datos$BQ_NUM[i] == 'NS/NC') {
    datos$BQ_NUM1[i] <-'NS/NC'
  } else {
    datos$BQ_NUM1[i] <-'4 o más'
  }
}

#Exportamos el conjunto de datos
write.csv2(datos, 'A:\\Aida\\Tessy\\Trabajo de Fin de
Master\\datos1.csv')

claboral <- read.sas7bdat('A:\\Aida\\Tessy\\Trabajo de Fin
de Mastercopia\\claboral.sas7bdat')

claboral$BCOTC142 <- claboral$BCOTC14

for (i in 1:nrow(claboral)){
  if (claboral$BCOTC142[i] == 'No\r\naplicable') {
    claboral$BCOTC142[i] <- 'No aplicable'
  } else if(claboral$BCOTC142[i] == 'No\r\r\naplicable') {
    claboral$BCOTC142[i] <-'No aplicable'
  }
}

write.csv2(claboral, 'A:\\Aida\\Tessy\\Trabajo de Fin de
Mastercopia\\claboral.csv')

```

```

#Bucle para ver las titulaciones por rama y clúster

union <- read.sas7bdat('A:\\Aida\\Tessy\\Trabajo de Fin de
Mastercopia\\union333.sas7bdat')

vrama <- c("Ingeniería y arquitectura", "Ciencias de la
salud", "Ciencias de la salud", "Ingeniería y arquitectura",
"Ciencias sociales y jurídicas", "Ingeniería y
arquitectura",
          "Ciencias de la salud", "Artes y humanidades",
"Ciencias sociales y jurídicas", "Ciencias")

vcluster <- c("ACAD_1", "ACAD_2", "ACAD_3", "ACAD_4",
"ACAD_5", "ACAD_6", "ACAD_7", "ACAD_8", "ACAD_9", "ACAD_10")

salidas <- list()
for (i in 1:10){
  salidas[[i]]
  <-
table(union$TITU[union$RAMA==vrama[i]&union$CLUSAC==vcluste
r[i]])
}

sort(salidas[[10]], decreasing=TRUE)[1:3]

#Arreglamos la variable edad
for (i in 1:nrow(union)){
  if (union$EDAD[i] == 'Menores\r\r\r\rnde 30 año') {
    union$edad1[i] <- 'Menores de 30 años'
  } else if(union$EDAD[i] == 'Menores\r\r\r\rnde 30 años') {
    union$edad1[i] <- 'Menores de 30 años'
  } else if(union$EDAD[i] == 'De 30 a 34\r\r\r\rnaños') {
    union$edad1[i] <- 'De 30 a 34 años'
  } else if(union$EDAD[i] == 'De 30 a 34\r\r\r\rnaños') {
    union$edad1[i] <- 'De 30 a 34 años'
  }
}

```

```

    } else if(union$EDAD[i] == 'De 35 y\r\r\nmás años') {
      union$edad1[i] <-'De 35 y más años'
    } else {
      union$edad1[i] <-'De 35 y más años'
    }
  }
}
for (i in 1:nrow(union)){
  if (union$TIC[i] == 'Usuario\r\r\r\nexperto') {
    union$TIC1[i] <- 'Usuario experto'
  } else if(union$TIC[i] == 'Usuario\r\r\r\nexperto') {
    union$TIC1[i] <-'Usuario experto'
  } else if(union$TIC[i] == 'Usuario\r\r\r\nde
nivel\r\r\r\nbásico') {
    union$TIC1[i] <-'Usuario de nivel básico'
  } else if(union$TIC[i] == 'Usuario\r\r\r\r\nde
nivel\r\r\r\r\nbásico') {
    union$TIC1[i] <-'Usuario de nivel básico'
  } else if(union$TIC[i] == 'Usuario\r\r\r\nde
nivel\r\r\r\navanzado') {
    union$TIC1[i] <-'Usuario del nivel avanzado'
  } else if(union$TIC[i] == 'Usuario\r\r\r\r\nde
nivel\r\r\r\r\navanza') {
    union$TIC1[i] <-'Usuario del nivel avanzado'
  } else {
    union$TIC1[i] <-'NS/NC'
  }
}
}
for (i in 1:nrow(union)){
  if (union$EST2_MS[i] == 'Artes y\r\r\r\r\nhumanidades'){
    union$EST2_MS2[i] <- 'Artes y humanidades'
  } else if(union$EST2_MS[i] == 'Artes y\r\r\r\r\nhumanidades')
{
    union$EST2_MS2[i] <-'Artes y humanidades'

```

```

    }else if(union$EST2_MS[i] == 'Ciencias\r\r\r\nsociales
y\r\r\r\njurídica') {
        union$EST2_MS2[i] <-'Ciencias sociales y jurídicas'
    }else if(union$EST2_MS[i] == 'Ingeniería
y\r\r\r\narquitectura') {
        union$EST2_MS2[i] <-'Ingeniería y arquitectura'
    } else if(union$EST2_MS[i] == 'Ciencias de\r\r\r\nla salud')
    {
        union$EST2_MS2[i] <-'Ciencias de la salud'
    } else if(union$EST2_MS[i] == 'Ciencias de\r\r\r\nla
salud') {
        union$EST2_MS2[i] <-'Ciencias de la salud'
    }else if(union$EST2_MS[i] == 'Ciencias\r\r\nsociales
y\r\r\njurídicas') {
        union$EST2_MS2[i] <-'Ciencias sociales y jurídicas'
    }else if(union$EST2_MS[i] == 'Ingeniería
y\r\r\narquitectura') {
        union$EST2_MS2[i] <-'Ingeniería y arquitectura'
    }else {
        union$EST2_MS2[i] <-'Ciencias'
    }
}

```

```

write.csv2(union, 'A:\\Aida\\Tessy\\Trabajo de Fin de
Mastercopia\\union444.csv')

```

A.5. Diagrama de SAS Enterprise Miner

A continuación, vemos el diagrama completo utilizado para el desarrollo del objetivo uno y la construcción del árbol de clasificación.

